The Journal of

APPLIED PSYCHOLOGY

EDITED BY

G. STANLEY HALL L. R. GEISSLER

And a Board of Co-operating Editors

VOLUME III

1919

WORCESTER, MASS.

FLORENCE CHANDLER, Publisher

Reprinted with the permission of the original publishers.

JOHNSON REPRINT CORPORATION

KRAUS REPRINT CORPORATION

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. BINGHAM, Carnegie Institute of Technology; JUNE E. DOWNEY, University of Wyoming; ROBERT H. GAULT, Northwestern University; WILLIAM HEALY, Baker Foundation, Juvenile Court, Boston; H. L. HOLLINGWORTH, Columbia University; JOSEPH JASTROW, University of Wisconsin; HERBERT S. LANGFELD, Harvard University; JAMES H. LEUBA, Bryn Mawr College; HENRY C. MCCOMAS, Princeton University; WALTER D. SCOTT, Northwestern University; C. E. SEASHORE, University of Iowa; DANIEL STARCH, University of Wisconsin; EDWARD K. STRONG, JR., George Peabody College for Teachers; L. M. TERMAN, Leland Stanford Jr. University; E. L. THORNDIKE, Columbia University; J. E. WALLACE WALLIN, Psycho-Educational Clinic, St. Louis; G. M. WHIPPLE, Carnegie Institute for Technology, Pittsburgh; HELEN THOMPSON WOOLLEY, Vocation Bureau, Cincinnati.

The scope of the JOURNAL OF APPLIED PSYCHOLOGY will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis (c) The influence of general environmental conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

COPYRIGHT, 1919, BY G. STANLEY HALL

TABLE OF CONTENTS

GRACE ARTHUR and HERBERT WOODROW
An *absolute* intelligence scale 118-137

W. V. BINGHAM
Army personnel work 1-12

CHESTER S. CARNEY
National conference on Americanization in industries . 269-276

J. CROSBY CHAPMAN
The learning curve in typewriting 252-268

J. CROSBY CHAPMAN and HERBERT A. TOOPS
A written trade test: Multiple choice method . . . 358-365

EDGAR A. DOLL
The average mental age of adults 317-328

JOSEPHINE CURTIS FOSTER
A case of intellectual development despite enforced seclusion 167-171

GEORGE W. FRASIER
A comparative study of the variability of boys and girls. 151-155

THOMAS H. HAINES and W. D. PARTLOW
Mental rating of juvenile dependents and delinquents in Alabama 291-309

V. A C. HENMON
Air service tests of aptitude for flying 103-109

C. M. KELLEY and F. L. WELLS
Briefer studies from the psychological laboratory of McLean Hospital 172-193

TRUMAN L. KELLEY
Principles underlying the classification of men . . 50-67

HARRY DEXTER KITSON
Comparison between two scales for the estimation of intelligence 310-316

FRANCES LOWELL
A group intelligence scale for primary grades . . . 215-247

W. D. PARTLOW and THOMAS H. HAINES
Mental rating of juvenile dependents and delinquents in Alabama 291-309

RICHARD H. PAYNTER, JR.
A psychological investigation of the likelihood of confusion between the words "coca-cola" and "chero-cola". 329-351

RUDOLF PINTNER
A non-language group intelligence test 199-214

RUDOLF PINTNER and HERBERT A. TOOPS
Educational differences among tradesmen 33-49

S. L. PRESSEY
The "efficiency" of a group scale of intelligence in prognosticating success and failure in junior high school . 381-385

CONTENTS

S. L. PRESSEY and L. W. PRESSEY
The practical "efficiency" of a group scale of intelligence. 68-80
"Cross-out" tests 138-150

S. L. PRESSEY and RUTH RALSTON
The relation of the general intelligence of school children to the occupation of their fathers 366-373

S. L. PRESSEY and I. M. SHIVELY
A practical information test for use with delinquents and illiterate adults 374-380

S. L. PRESSEY and G. F. TETER
A comparison of colored and white children by means of a group scale of intelligence 277-282

S. L. PRESSEY and J. B. THOMAS
A study of country children in (1) a good and (2) a poor farming district by means of a group scale of intelligence 283-286

RUTH RALSTON and S. L. PRESSEY
The relation of the general intelligence of school children to the occupation of their fathers 366-373

EDWARD S. ROBINSON
The analysis of trade ability 352-357

I. M. SHIVELY and S. L. PRESSEY
A practical information test for use with delinquents and illiterate adults 374-380

WALTER B. SWIFT
Mental attitudes in debate 156-166

G. F. TETER and S. L. PRESSEY
A comparison of colored and white children by means of a group scale of intelligence 277-282

J. B. THOMAS and S. L. PRESSEY
A study of country children in (1) a good and (2) a poor farming district by means of a group scale of intelligence 283-286

EDWARD L. THORNDIKE
A standardized group examination of intelligence independent of language 13-32

L. L. THURSTONE
Mental tests for prospective telegraphers 110-117
A standardized test for office clerks 248-251

HERBERT A. TOOPS and J. CROSBY CHAPMAN
A written trade test: Multiple choice method . . . 358-365

HERBERT A. TOOPS and RUDOLPH PINTNER
Educational differences among tradesmen 33-49

F. L. WELLS and C. M. KELLEY
Briefer studies from the psychological laboratory of McLean Hospital 172-193

HERBERT WOODROW and GRACE ARTHUR
An absolute intelligence scale 118-137

In Memoriam—JOHN WALLACE BAIRD . . . . . Frontispiece

A Symposium. The field of clinical psychology as an applied science 81-95

NOTES 195-198, 289-290, 393-396

BOOK REVIEWS 96-102, 194, 287-288, 386-392

In Memoriam

In the untimely death of Professor JOHN WALLACE BAIRD, of Clark University, the *Journal of Applied Psychology* has lost one of its founders and ardent supporters. Exhausting his energies in applying the facts of psychology to the task of rehabilitating the crippled soldiers returned from France for treatment at the Walter Reed Hospital, Professor Baird's already impaired health broke down so completely that the best medical skill at the Johns Hopkins Hospital could do no more than temporarily relieve the intense suffering and pain which he bore with heroic patience, cheerfully and without complaint to the last. He died on the evening of February 2nd in Baltimore and was buried in the town of his birth, St. Mary's, Ontario, Canada.

Born on the 21st of May, 1869, he studied at the Universities of Toronto, Leipzig, Wisconsin, and Cornell. At the last-mentioned institution he received the degree of Ph.D. in 1902 and served as assistant in psychology and later as Carnegie Research Assistant until 1904, when he was called to Johns Hopkins University as instructor. In 1906 he entered the teaching staff of the University of Illinois, and in

February, 1910, he came to Clark University to take charge of the psychological laboratory. He has conducted and supervised many important experimental investigations, especially in the field of the higher mental functions, such as memory and the thought-processes. The concrete problems involved in these subjects aroused in him a strong interest in the applications of psychology, as is shown not only in such studies as the legibility of printers' type and memory for absolute pitch, but also in his enthusiastic efforts in helping to launch the publication of the *Journal of Applied Psychology*. When finally the opportunity presented itself to make use of psychological principles in rehabilitating wounded soldiers and helping them to regain control over their crippled bodies, he applied himself to this task with an untiring devotion and self-renunciation until his own strength was exhausted.

His deep scholarship is recognized by colleagues and associates in requesting his services on many important missions and in his election to the presidency of the American Psychological Association. He was loyal to his friends and deeply devoted to his family, and all who came in contact with him will remember him as a true gentleman.

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. III MARCH, 1919 No. 1

ARMY PERSONNEL WORK.

WITH SOME IMPLICATIONS FOR EDUCATION AND INDUSTRY.

By Lieutenant Colonel W. V. BINGHAM, Carnegie Institute of Technology.
(Formerly Executive Secretary of the Committee on Classification of personnel in the Army)

Every practical problem of personnel in war, industry, business or education presents at least three phases; first, study of the job, that is, determination of the duties to be performed and specification of the qualifications necessary to perform those duties; second, inventory of the individual man to ascertain and record his qualifications; third, placement of the man in the job. In the Army this implies a system of personnel classification, a routine for making accurate requisitions, a procedure for locating quickly the specified personnel, and a machinery for effecting its assignment or transfer to the post where it is most needed.

It is the contention of the writer that the most fundamental phase of personnel work and yet the most neglected, is the first mentioned; namely, analysis and precise definition of duties and qualifications.

Consider that almost every company in the Signal Corps and the Air Service has one or more so-called "Master Signal Electricians." If a number of soldiers had to be trained to fill these various posts within a few weeks, what sort of recruits would you select for training? You would be at a loss until you knew the duties. In an Outpost Company of a Field Signal Battalion, a Master Signal Electrician is "responsible for the installation, maintenance and operation of small telephone offices, and for the proper functioning of

electrical apparatus such as buzzer-phones and short range radio apparatus." But in each Pigeon Company of 324 men there are also two Master Signal Electricians; one is " in charge of the training and instruction of the men as pigeoneers and is responsible for the care and condition of the pigeons." The other, " as loft master, is in charge of the breeding loft and the 90 mobile lofts, and is responsible for the proper movement of the latter. He calculates the needs of the lofts both as to pigeons and as to supplies." In the Telegraph Company you want neither a wire chief nor a racing pigeon fancier, but a telegraph engineer to " make studies and survey routes for construction of permanent telephone or telegraph lines;—and specify layout of telephone and telegraph offices and power plant for common battery telephone switchboards." The Supply Section of an Aero Supply Squadron requires as its Master Signal Electrician an airplane mechanic; other companies need a tinsmith, a coppersmith, an armorer, a toolmaker, a master painter, an instrument repairman, a construction foreman, a storage battery electrician, a truckmaster, all masquerading with the title, grade and pay of Master Signal Electrician. In the Fabric Workers Company this particular soldier must be a tailor, for his duties are " to supervise repairs and replacements to coverings for wings, fuselage and tail surface (in linen covered sections.)

Another common grade in the Army is that of " Wagoner." In some types of units the wagoners are the soldiers who have to repair the wagons, while in other units they are mere wagon drivers, and in still others they must drive automobiles. The " horseshoer " sometimes shoes horses, but in a motorized regiments the soldier with that title repairs the motor trucks. One hesitates to make any assignments of skilled personnel in default of some definition of duties. But when you have secured a statement of just what the soldier or officer has to do, you have taken a long stride toward judicious placement and training. Without such a definition you know neither whom to choose nor how to train, and gross wastage of precious time, money and life goes on with an extravagance approximated only by the blithe wastefulness of college class rooms.

When the mobilization of the National Army began in September 1917, and the nation poured into the cantonments its human wealth of trained artisans, teachers, business and professional men, laborers, farmers and shop hands, many of them illiterate or non-English speaking, the newly appointed Personnel Officers had extremely little information to guide

them in making judicious assignments of such varied skill and talent. The preparation of this information has demanded an exhaustive study of the entire Army organization to determine where ability of various kinds is required. Its effective utilization has necessitated the development and supervision of an Army personnel system to discover the occupational, educational and military qualifications possessed by the recruits and to insure their assignment to the proper units.

These tasks were assigned to the Committee on Classification of Personnel in the Army, created by Secretary Baker on August 5, 1917, as an instrument to increase the value of the Army's man-power through securing the most effective placement of each man.[1]

For this program an initial appropriation of $25,000.00 was made, and as the scope of the Committee's responsibilities grew, additional appropriations were approved until the total amounted to $851,650.00.

While it happened that the original membership of the Committee consisted almost wholly of psychologists, many industrial and business specialists in employing, classifying and assigning men were called upon to insure the successful prosecution of the work. This has included the following activities:

1. *Classification and Placement of Enlisted Men:*

Personnel offices have been establishd in all Army divisions, depot and training camps, coast defense stations, aviation fields, special training camps for Staff Corps and at other Army Posts. In these offices a special card system furnished accessible information as to the educational, occupational, and military qualifications of every man. With a minimum of clerical work this system selected 973,858 men for transfer largely into technical units in the Engineers, Aviation, Ordnance and other Staff Corps, and even more men for transfer within the divisions or camps. Sixteen civilian supervisors, directed by

[1] The Committee was organized with Walter Dill Scott as Director, E. L. Thorndike as Chairman, and the writer as Executive Secretary. The other members were J. R. Angell, R. C. Clothier, Raymond Dodge, H. L. Gardner, J. F. Shepard, E. K. Strong, Jr., L. M. Terman, J. B. Watson, and R. M. Yerkes. The places of five of these original members have since been filled by J. J. Coss, W. R. DeField, W. B. Hale, Winslow Russell, P. J. Reilly and J. J. Swan. A large force of able, devoted civilian and military associates of the Committee, in Washington, in the camps, and overseas, has made possible the realization of its plans.

the Committee, acted in an organizing and supervisory capacity in the field. Approximately 450 officers and 7,000 men were engaged in this personnel work. The number of soldiers interviewed by trained examiners and classified according to their best Army usefulness was in all, approximately three and a half million men.

2. *The Allotment Branch or Central Clearing Office*

of the Committee in Washington received reports on the numbers of skilled tradesmen found in each contingent of the draft, received and consolidated requisitions from the Staff Corps for specialists, and prorated these requisitions among the various camps according to their supply of necessary skilled men. On November 11, requisitions for roughly 600,000 men of designated qualifications had been filled here. Information was available at any moment for the Operations Division of the General Staff concerning the occupational qualifications of all the men in the several depot brigades, Army vocational schools, and similar sources of supply.

3. *Trade Specifications and Index of Occupations:*

Definitions of the many hundred different trades needed in our military establishment were prepared after exhaustive study, and were brought together in a book, "Army Trade Specifications." This index is an indispensable reference for Staff Corps and camp personnel officers in securing the skilled personnel needed.

4. *Tables of Occupational Needs and Personnel Specifications:*

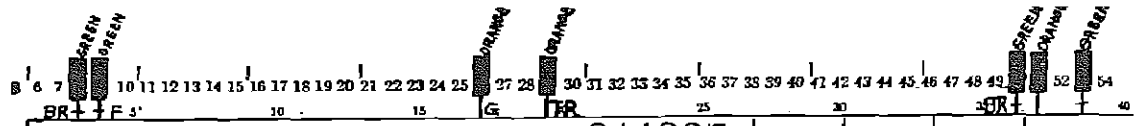
Tables were prepared showing in detail the needs for skilled workers in each sort of platoon, company, regiment or other unit. These tables were studied, criticised, and approved by Army units at the front in France, and later formed the basis for organizing quickly the newest divisions. Out of these Occupational Tables have developed the Personnel Specifications which have now been completed for the enlisted personnel of four hundred different kinds of organizations.

5. *Trade Tests:*

To increase the accuracy of selecting skilled workers among the enlisted men, a system of practical trade tests was devised, standardized and installed in twenty camps. At the time of the armistice about 130,000 men who claimed occupational skill had been trade tested.

5

ARMY PERSONNEL WORK



Form CCPT 1 11-1-18

SOLDIERS' QUALIFICATION CARD

Name (Print clearly) Kellogg (Last name) John (First name) Lee (Middle name) 811227 (Number) (Rank or Grade) (Rate or Grade) (Rank or Grade) W (Race)

Limited Service

Main Occupation: Timber & Concrete Foreman — Yrs. at it: 3½ — Wkly. wage: $ 45

Just what did you do? Boss, heavy timber work and concrete construction, bridges, mills warehouses, etc. — Symbol: 506R A, 50C (3) — Surgeon's O.K.: ✓

Firm worked for? A. J. Smith Construction Co.

Address of Firm? 528 - 11th St, Chicago

Dept., Shop, or Branch: Denver — Kind of Business: Gen. Construction

2d Best Occupation: Timber and Concrete Carpenter — Yrs. at it: 4 — Wkly. wage: $ 33

Just what did you do? All round experience as carpenter on rough construction & Concrete forms bridges, mills, warehouses, etc. — Symbol: 98R A, 8CO (3) — Surgeon's O.K.: No.

3d Best Occupation: Tracer (draftsman) — Yrs. at it: 1½ — Wkly. wage: $ 15

Just what did you do? Traced plans - mainly floor plans Building construction — Symbol: 29-7R (4) — Surgeon's O.K.: ✓

Additional Occupations, hobbies, etc. 2 yrs. as general helper to Bricklayer Can mix mortar & lay Brick, 16G-(5)

Army Specialty: Limited Service

Assigned by Personnel Officer as: Bridge Carpenter — Final Physical Classification: C-2 — Preferred: ✓

Age now: 30 — Married? m - 3 — Non-E — Illiterate — Intelligence: A — S.D.

Age at end of Schooling: 19 — Height: 5'-10 — Weight: 185 — Date of Interview: MAR 12 1918

TRADE TEST RATINGS

Symbol	Rating	Symbol	Rating	Symbol	Rating	Symbol	Rating
1 8-BR	E	2 26-G	A	5		7	
3 8-G	J	4		6		8	

Greatest Authority or Leadership: Boss, construction work, 50 men, 3½ yrs.

PRESENT ASSIGNMENT AND FUTURE CHANGES

	Company	Organization	Div'n or Corps	Date
1		Camp Hq.	2 M	MAR 12 1918
2				
3				
4				
5				

⅝ actual size of card

SCHOOLING	No. of Yrs.	Graduated Yes	Graduated No
Common School	8	✓	
High School	4	✓	
College or University			
Technical College			
Trade, Night, or Business School	1		

ABILITY TO CONVERSE IN:—

	French (2)	German (3)	Some other language (4)
Fairly Well	—	—	—
Perfectly	—	✓	

Institution — Subject of Specialization

Trade, Night, or Business School: Y.M.C.A. Drafting

Army Schools at	No. wks.	Course	A. J
1			
2			
3			

Previous Military Experience: Rank; Years of Service; Organization: Corpl. 1 yr. 9 Illinois National Guard.

Branch of Service or Special Duty Preferred: Engineers

Religious Preference? Presbyterian (answer requested but not required)

Member of What Fraternal, etc., Organizations? Elks, Carpenters Union

Talent for Furnishing Public Entertainment: Play guitar

ARMY SPECIALTIES (For Overseas Use Only)

1 Skipper	10 Rifle Grenadier	19 Ammunition Cannoneer
2 Pioneer	11 Rifleman	20 Gun Cannoneer
3 Stokes Mortar	12 Automatic Rifleman	21 Driver
4 1-pounder	13 Company Clerk	22 Instru. Man (Artillery)
5 Telephone	14 Runner	23 Agent of Communication (Artillery)
6 Radio	15 Machine Gunner	24 Lithographer
7 Visual Signalman	16 Horseshoer	25 Powderman
8 Wagoner	17 Packer	26 Rigger
9 Hand Bomber	18 Gunner (Artillery)	

¾ actual size of card

Birthplace of father: Lee, Mass. of mother: London, England (Give City and State of U. S. or Province and Name of Foreign Country)

of soldier: Chicago, Ill. How long in U. S. — yrs.

Citizen ☑ Taken out first papers ☐ Non-citizen ☐

From Local Board 17 (No.) Chicago (Town, City or County) Ill. (State)

No. 283 Red Ink No. 183 Date of Induction MAR 9 1918

PHYSICAL FINDINGS	Date	SUBSEQUENT CHANGES
1. Genito-Urinary		
2. Orthopedic 2nd degree flat feet		
3. Medical Mitral regurgitation, mild, compensated		
4. Surgical		
5. Nervous and Mental		
6. Eyes R 20/40 L 20/200 Cor R 20/40 L 20/40		
7. Ear		
8. Nose and Throat		
9. Teeth		
10. Convalescent from		
11. Remarks Has had foreman, similar work, not as workman.		Preliminary Physical Classification

REMARKS

Signature of soldier: John L Kellogg

Name of Camp: GRANT Initials of Interviewer: A.G.S.

6. *Personnel Work for Officers:*

Qualification Cards for officers, furnishing a record of occupational, educational and military experience and a rating by superior officers, were developed and put into use throughout the Army. These cards are filed in Washington, and duplicates filed in the custody of the Division Commanders for their own use in assigning their officers.

7. *Rating of Officers:*

A uniform system of rating officers was developed. This rating system was first installed in the Officers' Training Camps as an aid in selecting candidates for commissions. Later it was used in selecting candidates for Officers' Training Schools. Now its use is universal both in America and in France as a means of securing every three months, a rating on every officer as an aid in determining promotion, demotion, discharge, and appointments to the Officers' Reserve Corps.

8. *Commissioned Personnel Specifications:*

Definitions of duties and qualifications of no less than 500 different kinds of officers in the various arms and branches of the service have been prepared, after ascertaining the answer to the question, "Just what does he do?" These commissioned personnel specifications are for use in recruiting officer material, in selecting men for training as officers, and making assignments. Statistical studies have been made of the relative significance of age, education, civilian earnings, intelligence and other qualifications of officers in the different branches of the service.

9. *Cooperation with the Air Service:*

The methods of selecting aviators were investigated and checked by reference to the actual success or failure of the pilots. An improved system of tests for aviation candidates was introduced and a new program of examination and selection installed.

10. *Cooperation with the Provost Marshal General's Office:*

Plans for securing classificatory information regarding all registrants were submitted to the Provost Marshal General and were partially embodied in the draft questionnaire.

11. *Development Battalions:*

The Committee cooperated closely with the General Staff and the Surgeon General's Office in preparing and introducing the plans for segregating, sorting, training, and utilizing the partially fit.

12. *Cooperation with the Surgeon General's Office:*

Some assistance was given to the Psychology Section of the Surgeon General's Department in devising and administering the intelligence tests for enlisted men and officers. The psychologists in turn have tested 1,716,000 soldiers and furnished the personnel officers with their intelligence ratings to assist in making assignments, balancing units, and selecting men for special responsibilities.

13. *Cooperation with the Navy:*

At the time of the armistice, representatives of this Committee were assisting the Navy and the Marine Corps to prepare and install a complete personnel classification program to correspond with the Army system. In addition, one member of the Committee has done valuable work in refining methods of selecting and training men for special duties in the Navy, such as the work of the fire-control squad, the gun-pointer, the hydrophone listener, and the lookout.

14. *The War Service Exchange:*

This branch of the Committee was established January 18, 1918, to receive and classify applications of persons desiring to serve the Government and to refer them to the branches of the service needing them, and to cooperate with other agencies in locating and supplying men needed for special purposes by the various branches of the service. This organization relieved high officials of the War Department of the necessity of devoting valuable time to the interviewing of the many influential men who came to Washington to offer their services to the Government. It also cared for a total of about one hundred and ten thousand written proffers of service. It placed approximately ten thousand men, including many of superior attainments.

15. *Personnel Work in the American Expeditionary Forces:*

Members of the Committee studied the personnel needs of the A. E. F., and, with the cordial approval of General

Pershing, established there a personnel organization similar to that in America. The Officers' Qualification Cards have had their widest usefulness over-seas, in supplying replacements and in locating rare specialists in emergencies.

16. *British Experience:*

Detailed study was made of the working of the British personnel organization, which in some respects is far superior to ours. Special reports and exhibits obtained in London from the British War Office, covered their whole program of recruitment, classification, trade-testing, assignment and transfer; industrial furloughs; weekly consolidation and analysis of strength reports; and plans for demobilization.

This in outline is a picture of the personnel work, begun in the National Army cantonments with the arrival of the first trainloads of drafted men. Neither the civilians nor the Army Officers who initiated this development dreamed of the scope it would so rapidly assume or the share it would have in effecting the speedy organization of a well balanced Army, trained and ready for the critical hour in France.

The Committee on Classification of Personnel in the Army as such has disappeared. After fourteen months of service under the Adjutant General it was transferred to the General Staff and merged with the Central Personnel Branch, newly created to supervise the procurement, placement, transfer and promotion of officers throughout all branches of the Army. This means that centralized control of personnel work for both officers and soldiers is recognized and thoroughly established as an integral part of the United States Army organization.

Among the legacies left to the Army by the Committee, such as the qualification card, the index of occupations, the trade specifications and the standardized trade tests, not the least significant is the concept of personnel specifications with all that that term implies regarding analysis and definition of duties and formulation of requirements as to the physical, intellectual, educational, technical and personal qualifications required for the performance of those duties.

The concept of personnel specifications is quite as significant for industry and education as it is for war. The college student pursuing an electrical engineering course or a course in interior decoration, or the young woman who aims to be a

secretary or a social worker or a teacher of the arts and crafts, needs a clear definition of precisely the duties for which he or she is being trained.

Let the student have before him a full and explicit description of the *duties* of the job he must be able to hold down after graduation, and of the *qualifications* in the form of particular sorts of skill, technical knowledge, general abilities and personal traits which he must possess if he is going to make a success of that job. Let each instructor be in a position to speak with authority when challenged to justify every task he assigns as contributing directly to the development of one or more of these essential qualifications.

Let the factory foreman and the employment manager crystalize their formulations of the precise duties and requisite qualifications for every post. They will clear the way for a program of intelligent hiring, judicious placement, effective training, steady promotion and just remuneration.

SAMPLE PERSONNEL SPECIFICATIONS

HEADQUARTERS AND SUPPLY COMPANY

ANTI-AIRCRAFT SECTOR (ARMY ARTILLERY)

Army title set in bold face type followed on the same line by the corresponding civilian occupation, the number of men, the degree of skill and the occupational symbol.
On the next line appears the suggested substitute occupation.
Then follows the description of army duties.

---

2 Sergeants Major, Junior Grade:

1 Battalion Sergeant Major. Clerical worker, general — 1 Jour. 38g
Substitute: Clerk, railroad.

Chief clerk of sector office. Under direction of Adjutant, supervises the preparation and filing of reports, returns, correspondence, orders, and bulletins.

1 **Supply Sergeant Major**.....Stockkeeper — 1 Jour. 18s
Substitute: Merchant.

Calculates the needs of the sector in quartermaster, signal, engineer, and ordnance property; requisitions and issues same, including daily allowance of food and fuel, to the batteries; keeps property accounts with supply depots and batteries.

1 First Sergeant: Foreman with qualities of leadership and mechanical ability..... — 1 Uncl.

Agent of company commander in maintaining discipline; manages company under direction of company commander, arranging details in execution of orders and assigning men to various employments; required to act on own initiative in emergencies in absence of company officers; supervises the preparation of company reports, returns, and correspondence.

1 Electrician Sergeant, First Class..Electrician, general — 1 Jour. 10g
Substitute: Telephone troubleman, outside.

Has charge of communications system of the sector, exclusive of wireless. Maintains and repairs telephone system, locating "troubles" and supervising the work of wiremen and repairmen. Maintains other electrical equipment of sector, including searchlights.

1 Assistant Engineer...Gasoline engineman, stationary — 1 Jour.25s
Substitute: Auto repairman, general.

Has charge of power-generator system, gasoline engines, and motor transportation of the sector, directing the repairs necessary to their proper maintenance.

1 Radio Sergeant.....Wireless constructor — 1 Jour. 31wc
Substitute: Wireless operator.

Has charge of the wireless apparatus of the sector; constructing, maintaining and supervising the operations of the radio system, and the instruction of radio operators in the operation, care, and repair of wireless instruments and equipment.

SAMPLE PERSONNEL SPECIFICATION
(OFFICERS)

Name of Corps or Arm: Coast Artillery Corps

Subdivision or Organization: Battery, Motorized Gun or Howitzer

Official Designation and Rank of Officer: Battery Commander, Captain

I. Description of Duties.

Commands 4 lieutenants and 219 to 278 men. Responsible for preparation of company reports and records; for the proper equipment of his command and for the messing of the personnel. Responsible for property of the battery; guns or howitzers, motor trucks and tractors; infantry equipment of each soldier and other company supplies.

Responsible for the training and discipline of officers and men as soldiers and their instruction in loading and firing of guns or howitzers, machine guns and small arms and their proper care; in motor transportation, observation, signalling and telephonic communication; in earthwork construction, in gas defense, and in camouflage.

Responsible for the tactical movement of his battery (men, guns and equipment); tactical preparation of position for guns; orientation (by means of surveying); calculation of firing data (involving use of trigonometry and logarithms) and correction of data when firing and under fire.

When battery operates independently will assume duties of Major, requiring tactical judgment in reconnaisance, and initiative in direction and supervision of fire.

II. SPECIAL TECHNICAL QUALIFICATIONS.

Required: Working knowledge of mathematics through trigonometry and logarithms.

Desired: Knowledge of mechanics, electricity, and motor transportation.

III. FIRST CHOICE OF CIVILIAN OCCUPATION.

	Years of Experience
Civil Engineer	3—5

IV. OTHER CHOICES OF CIVILIAN OCCUPATION.

	Years of Experience
Electrical or mechanical engineer	3—5
Graduate of recognized technical school with business or professional experience	3—5
College graduate with business or professional experience	3—5

V. STAFF CORPS SCHOOL.

Will be required to attend school for three months.

VI. AGE LIMITS.

Possible Age Limit 25-45 years.

Best Age Limit 30-40 years.

VII. PHYSICAL REQUIREMENTS.

Officer must be qualified for full military service.

VIII. MINIMUM SCHOOLING.

Two years of college.

A STANDARDIZED GROUP EXAMINATION OF INTELLIGENCE INDEPENDENT OF LANGUAGE

By Edward L. Thorndike, Teachers College, Columbia University

I. Description of the Tests

The practicability of having one examiner measure the intelligence of a large group of persons at one time has been fully demonstrated by four years' experience of the Metropolitan Life Insurance Company with a test devised by the writer, by the work of McCall for the New York State Ventilation Commission, and lately on an extraordinarily wide scale by the experience of the United States Army with the admirable examinations devised by the Psychology Committee of the National Research Council.

Such group examinations reduce expense, increase uniformity of procedure, and, if extensive enough, give a more accurate determination of intellect than would be obtained by an hour's interview by the average of the most competent third of physicians, teachers or psychologists. They cannot, of course, do the work of a specially talented diagnostician of intellect, who adapts his examination to the peculiarities of each case and sees the bearings of minute features of each response. But not over one in ten of physicians, teachers and psychologists would claim such special talents. They cannot detect malingering or suggest the causes of the degree of dullness found. They are therefore only accessory features of a general mental examination, inventory, and diagnosis. But where the persons examined are doing their best to score well, and in so far as the question is simply "How intelligent is this man or boy?" they are, if well devised, substantially adequate. And where large numbers are to be examined, they are a useful first step in a general mental examination, inventory and diagnosis.

The first forms of group examinations of intelligence consisted chiefly of intellectual tasks with words and numbers and presupposed a considerable ability to read. They were thus unavailable for use with foreigners and illiterates, were unfair to those of limited opportunity to learn to read well, and in all cases measured intelligence with words and

numbers to the relative exclusion of intelligence in dealing with things and people.

The psychologists in the army therefore made up from the various non-verbal tests which had been devised and experimented with, by Norsworthy, Knox, Porteus, Dearborn, Kelley, Pintner and Patterson and others, a composite examination, the directions for which could be given by gesture and the execution of which required no acquaintance with spoken or written English. This formed the well-known Beta examination of the army.

The present examination is of the same general nature as the Beta examination, but differs from it in the following points:

Ten alternative forms of the examinations are provided in place of the single form, to reduce the harm from unfair tutoring.

Each form is more extensive.

In the score, speed counts relatively much less, the candidate being given time enough to exhaust his ability in most of the eight tests used.

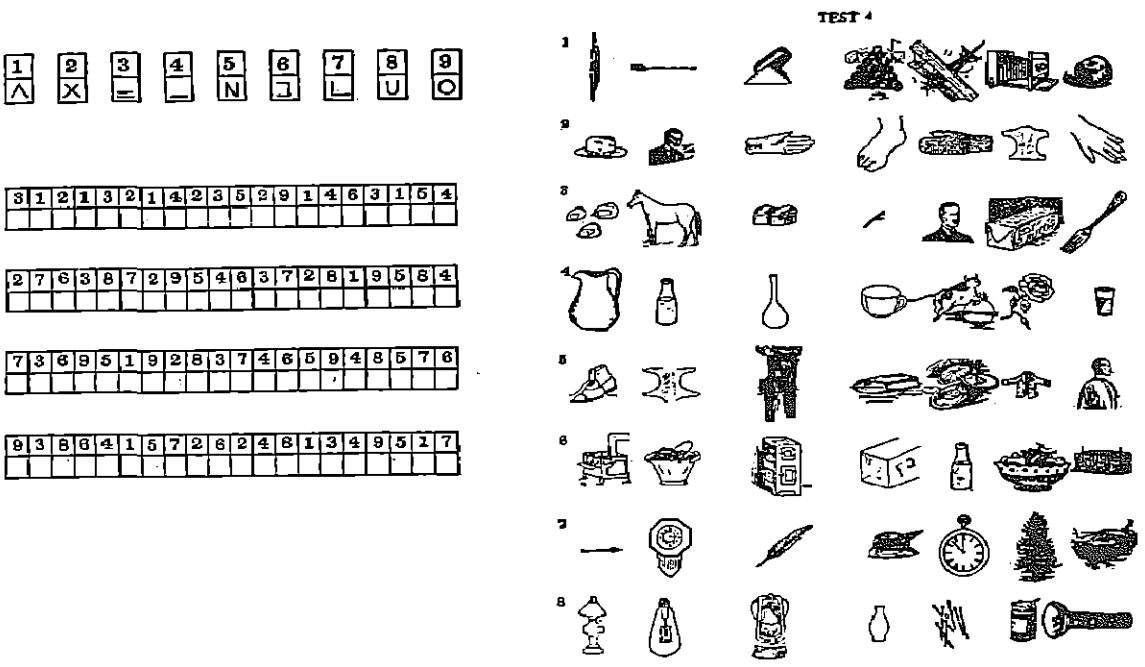
In the Beta examination the examiner was permitted to use speech, and to vary his treatment of groups somewhat. In the present examination all procedure is rigidly uniform.

In the Beta examination the effort was made to secure in the directions the maximum of comprehension of the general nature of each task, the score being intended to represent the candidate's ability to do harder and harder elements of an understood task and to do them quickly. In the present examination the effort is to make the directions identical for all and such as will be unvaried in their effect by the examiner's personality, zeal, dramatic ability, or experience in giving the examinations. Each of these plans has its peculiar advantages.

The content in the Beta examination was chosen primarily for high correlation with intellect as tested by the ordinary verbal forms of examination, such as the Stanford Binet or the Army Alpha examination. The Beta examination was intended for use as an alternative to Alpha. The content in the present examination was chosen primarily for high correlation with intellect as employed in managing both things and words. The examination is intended for use as a supplement to standard tests of verbal intelligence.

The following four types of tests used in the Beta examination are used also in the present examination: a test in substituting arbitrary symbols for digits; a test in drawing lines so

SERIES I. FRONT OF TEST-SHEET



SERIES L BACK OF TEST-SHEET

SAMPLE SHEET FOR SERIES I. FRONT

1	2	3	4	5
X	V	V	II	I

3	1	4	2	5	4

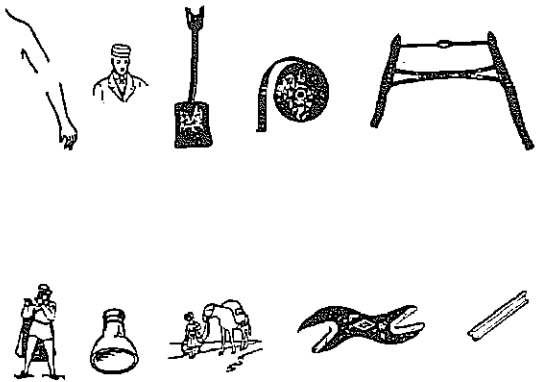
4	2	1	3	5	2

SAMPLE SHEET FOR SERIES I. BACK

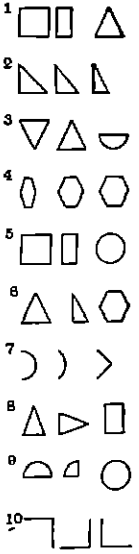
SERIES II
B

Write your name here
Write it very clearly

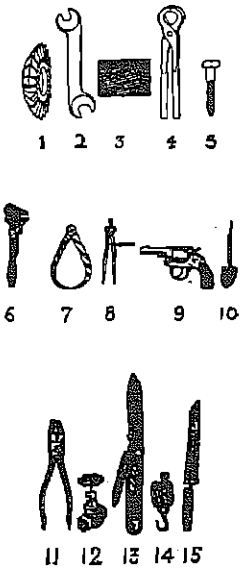
TEST 1



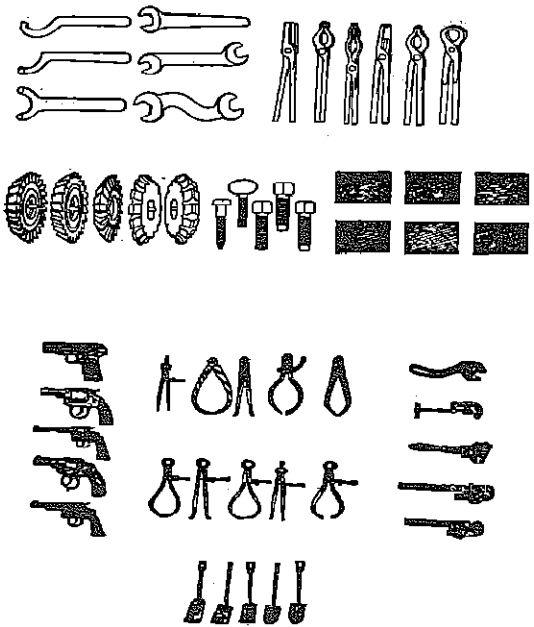
TEST 2

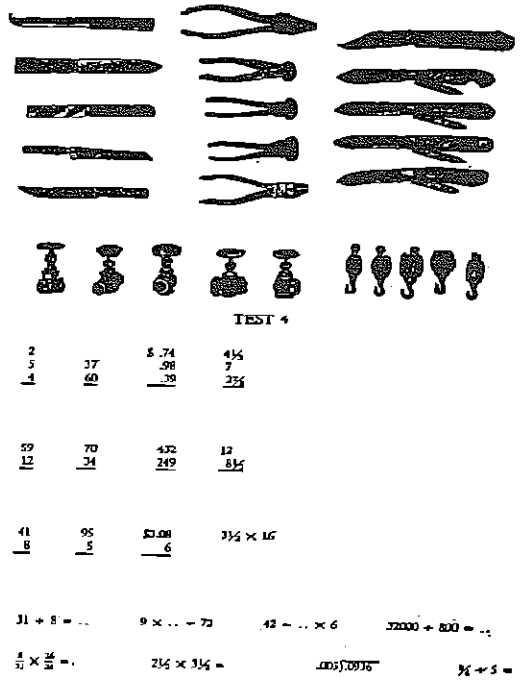


TEST 3



SERIES II. FRONT OF TEST-SHEET





SERIES II. BACK OF TEST-SHEET

SAMPLE SHEET FOR SERIES II. FRONT

SAMPLE SHEET FOR SERIES II. BACK

as to divide a surface into the parts shown beside it (the tasks are made rather harder, however); a test in completing pictures (the content here is however very different from that in the Healey, Pintner-Patterson, or Beta test, being made up largely of common material objects minus some essential part; a test in completing a rhythmic series of forms (the rhythms here are more varied, depend less on counting and are harder).

To these are added an analogies test with pictures,[1] a mixed spatial-relations test, a test of memory of objects, and a test in easy computation.

The material for an examination consists of two forms of Series I as shown on page 15-16, one form of series II shown on page 19-20, and the large sample sheets for I and II to be done by the examiner, shown on pages 17-18, 21-22.

The examination is given in three parts, I; I repeat, and II, the procedure being as follows: The candidates stand in a group of from 25 to 35 facing an easel on which is a large sheet, 36′ x 20′, like the *Sample Sheet for Series I, Tests 1 and 2*, shown below. The examiner says:

"We shall give you a sheet like this (holding up a Series I sheet). When you get it and are told to begin, do the work on it as fast as you can, but make no mistakes. If you come to something that you can't do, go ahead to the next thing. Do everything that you can. If you finish before time is called go back and make sure everything you did is right. Look at me. (If the group is large so that the two examiners are present, the instructions will be, 'Look at me or at the other man'). See what I do on this big sheet. When you get the sheet of paper, do the same as I did."[2]

He then fills the empty spaces in the five rows according to the pattern and draws lines cutting the squares into the parts shown, at a rate and in a manner prescribed for uniform use.[3]

He then turns the sheet over, displaying as its reverse the *Sample Sheet for Part I, Tests 3 and 4*. He writes X, V, etc.,

[1] This differs radically from that tried out and discarded by the Division of Psychology in the army. The test here is made very difficult, the four pictures from which choice is to be made all being relevant to one or more of the three which determine the choice, and the three "wrong" ones especially so. A right choice is not made by stupid people once out of four times nor even once out of ten times.

[2] These instructions may be spoken or displayed in printed form in any language desired.

[3] Detailed instructions for giving the tests are provided. The description here is abbreviated. Means to dispense entirely with language can be arranged if desired.

in the appropriate places, pointing to the blank spaces and to the key above after a prescribed method. In the last test of the four he marks the hat, cup, side-wheel steamer, and nest after certain instructive pointings.

The Part I test blanks are then distributed and 16 minutes are allowed for the work. At the end of 4 minutes the examiner says, "Even if you haven't finished Test 1, begin Test 2." At the end of 4 minutes more he says, "Even if you haven't finished Test 2, begin Test 3." Similarly at the end of 4 minutes more. Assistants see that these instructions are carried out by the candidates.

A different form of the same Series I is then used and with $2\frac{1}{2} + 2\frac{1}{2} + 2\frac{1}{2} + 2\frac{1}{2}$ minutes time, 10 minutes in all.

The candidates are then brought to the front of the room, the examiner directs their attention to a Part II sheet, and to the large Sample Sheet for Series II. Specimens of these appear on pages 19, 20 and 21, 22. The examiner fills in the missing parts, and draws the geometrical forms of Tests 1 and 2 of the sample sheet, making certain instructive pointings while doing so. He then does the work of Test 3, turning the large sample sheet over just as a candidate will have to do with the regular sheet, and finally does the computing of Test 4 on the large sample sheet.

The test sheets are then distributed and the men set at work. At the end of 4 minutes they are directd to begin on Test 2 if they have not already done so; at the end of 8 minutes they are directed to begin on Test 3 if they have not already done so. Similarly for Test 4 at the end of 12 minutes. At the end of 16 minutes all stop. The details for the administration of the examination will be reported elsewhere. One man can examine at one time as many individuals as can clearly see the large sample sheet, that is, from 30 to 40. An hour is required, of which 42 minutes are actual working time for the candidates. Any intelligent and careful person of the class of teachers, medical men, employment managers, social workers or the like can learn to give the examination adequately in an afternoon.

The results are quickly scored with the aid of keys. Tests I 1, I 3 and I 4, and II 3 and II 4 can be scored by any clerk. Tests I 2, II 1, and II 2 require either a moderate amount of judgment or a few hours of supervision by somebody of good judgment.

We have in this hour's examination obviously an extremely convenient instrument for measuring, and presenting in a numerical score, an individual's degree of success in meeting

certain demands on intelligence irrespective of language. If that score is significant of important facts about the individual, the examination is destined to be of value in schools, industry, business, social, philanthropic and legal work.

II. The Significance of the Examination Score

We shall consider first the significance of the score when the candidate takes it without having any previous acquaintance with it. In a later section we shall note any modifications or limitations to the significance of the score, supposing that the examination is public property (like the Stanford Binet) so that any person may study up for it as he may for a college entrance, or civil service, or immigration examination.

The examination, used under working conditions (i. e., in accordance with the detailed instructions prepared, by persons who had spent an afternoon in studying them and it, and a half-day in practice in testing school classes) yielded the following median scores for groups of increasing intelligence:

5 for 133 individuals 19 years old or over confined in institutions for idiots, imbeciles and the feeble-minded, and representing roughly the top tenth of their inmates in respect to intelligence. Less than 1 in 10 scores above 40.

95 for 171 sixth-grade boys largely of parents from the middle economic and social third of a large city.

115 for 190 sixth-grade boys largely of parents from the upper economic and social third of a large city.

170 for 169 fourth-year public high-school boys selected for continuance in school largely by ability. Only one in ten scores below 140.

210 for 104 superior adults, men commissioned or about to be commissioned in the United States Army. Only one in twenty-five scores below 140.

The distribution of all the individual scores for these groups, combining all the sixth-grade cases, is as shown in Table 1. It should be noted that the three individuals of the asylum cases who scored above 80 are described as follows by the authorities of the institution:

42 years old, understands most of the farm work, cares for horses and cows, drives team and plows, handles tools well. For quite a number of years has worked with engineers and steam fitters. Has done firing in boiler room under supervision. Enjoys and has a fair working knowledge of steam-fitting and ordinary electrical repairs.

TABLE I.

FREQUENCIES OF EACH SCORE IN THE EXAMINATION IN THE CASE OF UPPER-LEVEL FEEBLE MINDED IN INSTITUTIONS, 6TH GRADE BOYS, 4TH-YEAR HIGH SCHOOL STUDENTS AND SUPERIOR ADULTS. IN PERCENTS.

	0	1–9	10–19	20–29	30–	40–	50–	60–	70–	80–	90–	100–	110	120	130	140	150	160	170	180	190	200	210	220	230–239	240–249	250–259
F. M's. in Institutions.....	34	37	13	8	7	1¾	1¾	2	1¾		¾	¾															
6th Grade.....		¼		1	2	3	3½	7	7½	10	11	10	9	11	7½	10	4	2	1¼	¼							
4th Yr. H. S....								½			½	2	½	1	5	11	5	5	12	16	8	6½	3	2	2		
Superior Adults													1	1	2		3	5½	4	13¼	7¼	11½	13½	13¾	12¼	8½	3

19 years old, reads and writes, does a little number work, has very good memory. Does wood-work well. Makes brooms, brushes with little supervision. Is quite deaf. Has worked for several years with the carpenter, doing repair work in the manual training department. Usually his work is well done. Is capable of doing high-class work both in this and about the farm.

19 years old. No personal report was made, but he is recorded as testing 9.9 mental age by the Binet test, and can read and write.

In the case of the feeble-minded confined at one of the two institutions, data were available for the Binet mental age, the grade reached in school and the ability shown in doing work in the institution and its farm colony. Scores were given to each of the 79 individuals in the form of the sum of credits obtained by counting:

B

0 for Binet age 6 or below
2 " " " 7
4 " " " 8
7 " " " 9
10 " " " 10
14 " " " 11
20 " " " 12

Sc.

0 for grade 2 or below reached
3 " " 3
5 " " 4
8 " " 5
12 " " 6

Wo.

1 for 6 or 7 attained as the sum of 5 estimates of the work report
2 for 8 or 9 " " " " "
3 for 10 or 11 " " " " "
and so on, up to
10 for 24 or 25 " " " "

Te

0.2 X the test score.

We have correlations (by Sheppard's formula) as follows, showing that the test score is as good an index of intelligence in these cases as the Binet:

Binet with	B + Sc + Wo + Te	.79
Test "	B + Sc + Wo + Te	.93
Binet "	Sc + W	.77
Test "	Sc + W	.77
Binet "	B + Sc + Wo	.91
Test "	B + Sc + Wo	.84
Binet "	Sc + Wo + Te	.75
Test "	Sc + Wo + Te	.97

The authorities of the institution also reported that the rank-order of intellect obtained by the test tallied very closely with the estimates which they had formed.

A more exact result than the above could be obtained by assigning weights precisely so as to give the Binet and the Test Score absolutely equal weight in determining the composites. In the above Test Score has somewhat more weight, the distributions being:

Credits	Binet	Test Score	Credits	Binet	Test Score
0	14	23	11		
1		10	12		1
2	24	12	13		4
3		1	14		1
4	25	6	15		1
5		3	16		
6		7	17		
7	11	2	18		
8		1	19		2
9		1	20	1	1
10	4	2	21		1
					and 1 at 30

However, Sheppard's formula being used, the discrepancy will not amount to much; and it does not seem worth while to expend labor in the computation. The test-score is at least substantially as significant as the Binet Score.

In the case of 196 privates and non-commissioned officers who were ranked by a consensus of their officers and fellow soldiers for intelligence, the correlations were .23, .33, .43, .50, .53, and .58 for the six groups, with a median of .47 $\pm$ .05. This correlation is almost the same as that obtained between the consensus measure of intelligence and the regular army test Alpha in the case of sixty of these men who were tested with Alpha. The correlation with the real intelligence possessed by these men would presumably be higher, the amount of intellect imputed by the consensus being itself in each case an imperfect measure of the reality.

The test-score thus compares favorably with the score from the standard instruments for measuring intelligence, the Binet and the Alpha, in respect to the prophecy of intelligence given by it in the case of adults of low grade and adults in general. It is not suitable for distinguishing amongst adults of very high grade. Nor, in my opinion, is the Binet or the Alpha.

It has an advantage over the Binet, aside from its applicability to groups and to non-English speaking individuals. It is almost indefinitely extensible to alternate forms of equal difficulty so that a measure of any desired degree of reliability may be obtained, and measures of the effect of growth, training, fatigue, drugs or any other factor may be conveniently obtained. Ten forms are already prepared.

It has an advantage over Alpha aside from its non-verbal content, in that it distributes the cases of low intellect much more widely.

We have not been able to check the results against those from the Beta test, as yet. Presumably the two will do much the same work, the choice being primarily for convenience; though the lessening of the weight attached to speed seems to the writer theoretically and practically sound.

The test then seems destined to be useful alone, or with other "performance" tests, in the case of immigrants, illiterates, and all groups characterized by the absence of opportunity to learn English (or any other language) or great inequalities in such opportunity. It will be useful as an alternative for, or supplement to, the Binet, Alpha and similar tests in the case of those who can understand and use English.

The test scores may be transmuted into Alpha scores, or Stanford Binet Mental Ages, or standard points on the scale for sixth grade pupils, by Table II,[4] which is derived from the scores of 163 individuals tested with both this test and Alpha, 134 individuals tested with both this test and the Binet, and over a thousand individuals tested with this test alone.

These equivalents and other items entered in this table will give an adequate notion of the significance of any given score for any given person in respect to educational or vocational or general social promise, so far as these depend on the qualities measured by the test.

The scores of this table and all the results so far stated are for a first trial with the test by a person who has never been coached with a view to improving his score in this or a similar test. This introduces us to one of the chief difficulties of this and all other present instruments for the measurement of intelligence.

Improvement is made in score quite innocently in a second trial though the individual has never thought of the test in the meantime, by reason of the general familiarity with the tasks, adaptation to them, and whatever other special advantages a preliminary trial gives. In the Binet this averages about 3 mo. of mental age; in the Alpha the median gain is about 15 points (unweighted system) for men of A and B grade; in the present test the median gain is 10 points for sixth-grade school boys and average adult men, and 14 points for superior adults. This matter can be neglected so long as the tests are little known and used only for diagnosis and surveys. As they become widely known and are used to

[4] Table II is an abstract from a more extensive and detailed table.

TABLE II.

Test Score	Equivalent Alpha Score	Equivalent Stanford Mental Age in F. M.'s.	6th grade Boys	Stanford Mental Age Estimated
2		7½		
5	Median of 133 high grade.		F. M's. in institutions	
10		8		
25		8½		
37			—2.0 S.D.	
40	40			
45	42			9 yr. 6 mo.
50	44			
54			—1.5 S.D.	
55	47			
60	49			
65	51			10 yr. 6 mo.
70	53			
71			—1.0 S.D.	
75	55			
80	57			
85	59			11 yr. 6 mo.
87			—0.5 S.D.	
90	61			
95	63			
100	Median 12 yr. 0 mo. old boy		0	
105	67			
110	69			
115	70			
120	72		+0.5 S.D.	
122				
125	74			
130	76			
135	78			
139			+1.0 S.D.	
140	82			
145	86			
150	90			
155	93			
156			+1.5 S.D.	
160	97			
165	101			
170	Median 4th year H. S. boys			
172	106		+2.0 S.D.	
175	109			
180	114			
185	117			
190	120			
195	123			
200	126			
205	129	Superior Adult		
210	132			
215	135			
220	138			
225	141			
230	145			

measure progress or the effect of various factors, it must be allowed for, and tables for the significance of second-trial scores, third-trial scores and so on must be computed. The gain of a third trial over a second may be expected to be much less than that of a second over a first, roughly about .4 as much.

Improvement may be made in score by tuition over what would have been made without it. This is obvious to common sense. The tuition may be multiform, from general

coaching in taking tests to special drill in drawing a diamond or *counting* backward or defining words or learning opposites or understanding what is required in the analogies test or solving particular arithmetical problems. Experiment can determine the amount of influence of any defined amount of such coaching, and various measures can be taken to prevent illegitimate coaching from injuring the significance of a test score. Some of these are the provision of many alternative forms, the changing of test elements so that special coaching will mislead as much as it helps, the provision of a preliminary trial to put those who have not been coached more nearly on a level with those who have, so far as understanding the general nature of the test goes, and the separate computation of scores from the more coachable and the less coachable tests of the series.

In proportion as a high score in this test is used to confer any advantage upon an individual the procedure will have to be made adequate to prevent or at least detect such special coaching. This is even more true, however, of other standard examinations for intelligence, since they lack the protection of many alternative forms. Much such work must be done in the future.[5]

It should be noted that for many purposes the energy and ambition and industry of a candidate who studies ten or a dozen alternative forms of a test like this, so as to score high with whichever one confronts him at the examination, is a practical equivalent of the ability which enables him to score equally high without study. Such might be the case in many examinations for employment.

For cases where the test is to be used to measure progress or the effect of some factor, the reliability of the score is a matter of importance. The question as to how closely the score from one first trial of the test will represent the average score from an infinite number of first trials is of course essentially unanswerable. There is, however, a means of obtaining an approximate answer by the formula P. E. $= \sqrt{2}$ (median variation of difference between first-trial score and second-trial score).

According to this the probable error of a single determination of the test is:

[5] These matters have been more fully discussed by the writer in Tests of Intelligence: Reliability, Significance, Susceptibility to Special Training and Adaption to the General Nature of the Task, in *School and Society*, of Feb. 15, 1919.

14 for superior adults with a median score of 210
8 for adults somewhat above average with a median score of 140
8 for 6th grade pupils with a median score of 95

These results are based on repeated tests of 54, 46 and 143 individuals respectively. These figures are too high since they are a product of other causes of variability than the differences in the same ability in the same individual from one day to another. It is therefore conservative to estimate the Probable Error of a first time score as not over 6 or 7% of the score, for scores from 70 up.

The unreliability of a single second trial as a representative of the theoretical average of an infinite number of second trials, and similarly for a third or for a fourth trial, will presumably be less than the estimates given above for a first trial. Certain causes of variability are removed by acquaintance with the general nature of the test. We may, I think, expect that if two trials are given as practice tests and then a third and fourth trial as tests of morning versus afternoon, or of ability at age 15 versus ability at age 16, or of alcohol versus no alcohol, or the like, the probable error of the second score will be well under 6% of the score itself.[6]

[6] These tests may be obtained from the Bureau of Publications, Teachers College, Columbia University, together with the large sample display sheets and the standard instructions for giving and scoring the tests.

EDUCATIONAL DIFFERENCES AMONG TRADESMEN

HERBERT A. TOOPS and RUDOLPH PINTNER, Ohio State University.

In no other country and at no other time has there ever been such emphasis laid upon the necessity for a sound education in order to make good in the business of life. We believe in education. We know that good schooling is an asset to any man in any walk of life. The war has brought the fact home to us as never before. And the people of this country believe in it so much that they are willing to see their government spend thousands of dollars in order to allow some of the youth of the country to continue their college training at the country's expense.

That there is a relationship between the amount of education and the kind of position a man occupies is obviously true, despite individual exceptions. Between the trade and the profession there is a very marked difference. Does the same difference occur among the trades themselves? If so, what do we actually find to be the educational achievement of men in the various trades and how do the trades rank among themselves in regard to the educational attainments of the men?

The writers have had the opportunity of personally ascertaining the amount of education obtained by a great many tradesmen in a number of trades. The results and the conclusions derived are here presented, not as an exhaustive survey of even a single trade, but as indicative of some general existing tendencies.

The tradesmen questioned were employed in various shops of a large manufacturing city. They came from both large and small shops. In some cases practically every man of the trade in the city was questioned; in other cases only a liberal sampling was taken. The men in all the trades here considered are probably representative of the general type of man in each trade as found in this city.

In addition to the grade at which each man left school, there was determined as well the status of the man in his trade, whether apprentice, (A), journeyman, (J), or expert, (E). In rather a general way the definition of each status or class was as follows:

Apprentice—A beginner or learner in the trade, determined by length of experience in the trade, kind of shop, and the foreman's estimate of the man. In no trade was more than five years experience necessary as an apprenticeship stage for ordinary individuals.

Journeyman—A man past the apprenticeship period but not yet an expert. A practical criterion for a journeyman was that he was incapable of working without directions or incapable of directing others. A man with the requisite minimum of experience for an expert but not so classed by his foreman, was classed as a journeyman.

Expert—A man past a definite minimum of experience and classed as an expert by his foreman. A foreman of a shop was nearly always classed as an expert. A man capable of acting as a foreman or capable of working alone on difficult jobs without direction from others was also ordinarily classed as an expert.

It can be readily seen that there is no sharp dividing line between these classes, which are only arbitrary and artificial classifications. These three classes are, however, commonly recognized in the trades. In many cases it was difficult to decide in which status to put a given man. In such cases all possible evidence was taken into account and the man classified according to our best judgment based on this evidence. Since the dividing lines between the classes are only arbitrary divisions we find not only much overlapping of trade ability, but also much overlapping in any trait in which the classes might be measured, and therefore, in their educational attainments.

Many of the trades are similar in scope, as for instance, lathehand and toolmaker. Here we should expect great similarity in the education of the men in these trades. Other trades may be found which are very dissimilar, as horseshoeing and electrical instrument repairing. Between such trades we expect striking educational differences to occur.

Reports from 924 tradesmen are here given. These are distributed among thirty different trades. The blacksmiths were machine shop blacksmiths doing hand forging. The telephone linemen included a small number of telegraph linemen doing the same kind of work. The carpenters were house carpenters. Bench-hand is the name given to the trade of bench assembler and fitter. Powerhouse (A. C. and D. C.) were switchboard tenders on these respective kinds of switchboards. Electrical instrument repairers were ammeter and voltmeter repairers. General auto repair is but another name for general garage man. The ammonia pipefitters were re-

frigerating pipefitters. Storage battery men had to do with repairing automobile storage batteries. The switchboard repairers were telephone switchboard repairers at telephone stations. The carburetor men were general garage men who

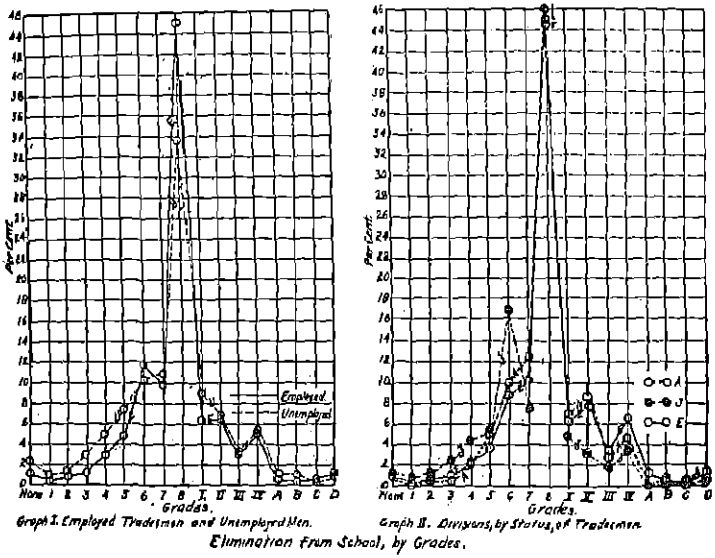
TABLE I. DISTRIBUTION OF TRADESMEN, IN NUMBERS AND PER CENT, BY GRADE AT LEAVING SCHOOL AND BY TRADE STATUS

Grade	Numbers				Per Cent.				
	Employed				9,506	Employed			
	Total	A	J	E	Unem.	Total	A	J	E
Column	1	2	3	4	5	6	7	8	9
None	9	2	4	3	2.1	1.0	0.8	1.3	0.8
1	3	0	2	1	0.8	0.3	0.0	0.6	0.3
2	7	1	4	2	1.3	0.8	0.4	1.3	0.6
3	11	1	7	3	2.7	1.2	0.4	2.3	0.8
4	26	5	13	8	4.8	2.8	2.0	4.2	2.2
5	43	13	16	14	7.3	4.7	5.1	5.2	3.9
6	109	25	52	32	10.2	11.8	9.9	16.9	8.8
7	91	31	23	37	10.6	9.8	12.3	7.5	10.2
8	417	113	142	162	33.8	45.1	44.9	45.9	44.6
I H. S.	56	18	15	23	8.8	6.1	7.1	4.9	6.3
II H. S.	59	21	10	28	6.6	6.4	8.3	3.2	7.7
III H. S.	25	7	6	12	3.0	2.7	2.8	2.0	3.3
IV H. S.	46	11	11	24	5.2	5.0	4.4	3.5	6.6
A College	5	..	..	5	0.8	0.5	...	...	1.4
B College	3	2	..	1	0.8	0.3	0.8	...	0.3
C College	4	..	2	2	0.3	0.4	...	0.6	0.5
D College	10	2	2	6	0.8	1.1	0.8	0.6	1.7
Total	924	252	309	363	99.9	100.0	100.0	100.0	100.0
Part H. S. or better education					26.4	22.5	24.2	14.9	27.8
Part College or better education					2.8	2.4	1.6	1.3	3.9
Median Grade at leaving school					8.30	8.39	8.43	8.24	8.51
Median Grade of Jour. + Experts					..	..	..	8.38	

were specialists on the carburetor work of gasoline motors for automobiles.

The number and percentage distributions of the tradesmen by grade at leaving school and by trade status, are shown in Table I. For comparison with the results for all 924 men (column 6) are also shown the percentage results from 9,506

unemployed men (column (5) previously reported.[1] The median grade of the employed men is found to be only 0.09 grade higher than the unemployed median. This might seem to point to no very great difference between the two groups. We have also previously pointed out the great stability of the medians of such large educational groups, and the significance of small changes in the median as well as the necessity of basing our judgment upon the actual distribution rather than the central tendency.[2] By plotting the percentages of columns 5 and 6, we obtain the curves of Graph I. We note the great



Graph I. Employed Tradesmen and Unemployed Men. Graph II. Divisions, by Status, of Tradesmen.

Elimination from School, by Grades.

inferiority of the unemployed group up to about the middle of the fifth grade shown by the large percentages of unemployed in the lower grades as compared with the percentage of employed in these same grades. The ratio of the two series of percentages is the real comparative measure of the two groups. Almost twice as many unemployed men leave school at each grade up to grade five as of employed men. In

[1] Toops, H. A. and Pintner, R.—Mentality in Its Relation to Elemination from School. *School and Society*, Vol. VII, Nos. 174-175, pp. 507-510, 534-539, April-May, 1918.

[2] Toops, H. A. and Pintner, R.—Variability of the Education of Unemployed Men. *Journal of Applied Psychology*, Vol. II, No. 3, Sept. 1918, pp. 207-218.

consequence of this, there is a much larger percentage of employed men than unemployed leaving at the eighth grade. In terms of ratios of percentages, the differences of the two groups at the eighth grade is much less significant than the difference between the two groups at the third grade. In the higher grades the differences are not so striking. Curiously enough, the "psychology of completion"[9] is much more marked at the seventh grade in the case of the employed men than the unemployed men, as we might expect.

By plotting the percentage distributions of the trade classes, we find rather marked educational differences as shown in Graph II. The amount of school elimination for apprentices is very similar to that of experts, and there is much overlapping of these two curves. In the case of the journeymen, we find a marked differentiation between this class and the other two classes. The journeymen have distinctly more elimination in the lower grades and much less in the higher grades, thus proving them to be, as a class, quite inferior to experts in the amount of education attained. This fact is also indicated by their median, 8.24, which is the lowest of all three medians as shown on Table II. The median of the apprentices is lower than that of the expert, as we should expect, if education has anything to do with trade success. The apprentices, by reason of being beginners in the trade, must contain both potential experts and also those who, even with the requisite amount of years of experience in the trade will not become experts. The effect of the "psychology of completion" is quite marked in the case of the journeymen only. Instead of this, we notice in the elimination curve an entirely new phenomenon, namely, a sudden rise in elimination in both apprentices and experts at the II grade of High School. Such irregularities in the curve may of course possibly be due to the comparative smallness of the numbers involved; the possible significance of these, should they prove to be real phenomena peculiar to these classes, will be discussed later. Experts and apprentices differ mainly from the journeymen in the relatively low proportion of poorly educated men and the relatively high proportion of better educated men with respect to the corresponding journeyman percentages.

The median grades by trades and the number of cases on which these are based are shown in Table II. The variation of the medians is not great, the range being only about three grades. In the division by trade status, in only one class, Journeymen, is the average deviation of the individual trade

[9] Cf. *Op. cit.* in Footnote 1.

TABLE II. DISTRIBUTION OF CASES BY TRADES; AND MEDIAN GRADES BY TRADE CLASSES AND BY TRADES

Trade	Rank Value	Number of Cases Total	A	J	E	Median Grade* A	J	E	Total
Horseshoer	1	9	2	6	1	...	....	...	7.0
Cobbler	2	21	5	5	11	...	....	7.5	7.3
Blacksmith(Shop)	3	18	..	9	9	....	6.7	8.3	7.5
Bricklayer	4	53	12	20	21	8.3	7.9	8.0	8.0
Lineman(Teleph)	5.5	54	14	12	28	8.4	8.3	8.1	8.2
Carpenter	5.5	42	14	19	9	8.6	6.5	8.4	8.2
Locomotive Repair	9.5	33	..	17	16	....	6.8	8.7	8.3
Harness Maker	9.5	10	..	2	8	....	....	8.3	8.3
Bench Hand	9.5	43	15	15	13	8.4	8.3	8.5	8.3
Steamfitter	9.5	27	16	1	10	8.2	...	8.4	8.3
AutoTruckMech.	9.5	45	12	18	15	8.1	8.2	8.8	8.3
Typewriter Repair	9.5	30	4	10	16	....	8.4	8.4	8.3
Sturc. Steel Erec.	14.5	20	1	6	13	....	....	8.4	8.4
Sheetmetal Work	14.5	16	3	6	7	....	....	8.6	8.4
Powerhouse(A.C.)	14.5	47	13	16	18	8.2	8.3	8.8	8.4
Powerhouse(D.C.)	14.5	38	12	16	10	8.4	8.1	8.9	8.4
Toolmaker	17	50	24	14	12	8.6	8.2	8.8	8.5
Plumber	20.5	47	12	15	20	8.8	8.5	8.6	8.6
Vulcanizer	20.5	18	4	2	12	...	....	8.6	8.6
Motor Repairer	20.5	38	13	14	11	I.0	8.4	8.5	8.6
Elec. Inst. Rep'r.	20.5	20	13	3	4	8.3	....	....	8.6
Inside Wireman	20.5	43	11	19	13	8.5	8.4	II.0	8.6
Gen. Auto Repair	20.5	40	14	11	15	8.5	8.5	8.9	8.6
Patternmaker	24.5	40	18	14	8	8.8	8.8	8.9	8.7
Ammonia Pipef'r.	24.5	17	3	10	4	...	8.9	....	8.7
Lathe Hand	26.5	25	8	8	9	II.2	8.6	II.0	8.8
Motorcycle Rep'r	26.5	17	..	7	10	....	8.8	8.9	8.8
Storage Batym'n.	28	19	..	4	15	....	....	I.5	I.5
Switchboard Repairer	29	31	9	8	14	III.0	I.3	8.9	I.6
Carburetor Man.	30	9	..	..	9	....	....	II.0	II.0
Total[@]—All Trades		920	252	307	361	8.43	8.24	8.51	8.39
		Average Deviation				0.43	0.53	0.43	0.38

* Median not calculated for less than 7 cases.
@ One trade with only 4 cases not included in this table.

medians from the median of all the men in the status, irrespective of trade, more than one-half a grade. We note also that the deviation of the total 924 men by trades from the median of the whole is less than any of the classes. This is one more indication of the remarkable stability of the central

tendency of education when large groups of men from many trades are considered; or in other words, the more nearly we approach an unselected sampling of the total population.

One may quite reasonably ask whether these results are based on a sufficiently large enough number of cases, and whether there actually exist such basic differences in the schooling of different tradesmen, as to consider the median grade a reliable index. After the results of the thirty-one trades had been tabulated, similar results were secured for six trades in a city in a different part of the country. Two of these trades, cook and brass melter, were not included in the 30 trades studied, and data were also secured for these at the first city, designated as City A. The resulting medians, given in Table III, suggest that even in widely different parts of the country, the medians by trades are not apt to vary on the average more than 0.3 of a grade. Those trades were selected

TABLE III. RELIABILITY OF TRADE MEDIAN GRADES

Trade	City A		City B	
	Cases	Median	Cases	Median
Horseshoer	9	7.0	12	7.4
Carburetor	9	11.0	21	11.3
Blacksmith	18	7.5	13	7.3
Motor Repairer	38	8.6	17	8.5
Cook*	20	8.1	37	8.0
Brass Melter*	21	7.5	30	6.9

* Data received too late to be included in Tables I, II, and IV.

which showed graduated differences in median from lowest to highest in order to test out whether the same significant range would be found at City B.

If we examine carefully the medians of Table II in the light of our knowledge of the trades, we may draw several conclusions therefrom.

In verification of the general results that journeymen are more poorly educated than either apprentices or experts as shown by the medians of the classes, we find that in 9 or three-fifths of the trades for which the A, J and E medians are calculated, journeyman medians are the lowest of the three. In 17 or 85 per cent. of the trades in which both J. and E medians are calculated, the journeymen tradesmen are less well educated than the experts.

In those trades in which the expert has a radically different line of work from the apprentice, the experts are much better educated. Examples are the auto truck mechanics in which the expert usually is a foreman of a general auto repair business including touring cars, and the two powerhouse trades in which the expert is usually a dynamo tender as well as a switchboard tender.

In trades involving great personal discomfort we are likely to find the general education low. Examples are the trades of brass melter, cook, blacksmith and horseshoer. In each of these trades we find a poor general education, and in each the personal discomfort due to heat is well marked. We believe that men are attracted toward the cleaner, more comfortable and better paid trades, and that the better educated are somehow able to secure the majority of these jobs thus forcing the less well educated to take the more undesirable jobs. In such trades, when work is plentiful, we find only the poorest type of apprentices entering the trade. Not one of the apprentice cooks or brass melters in City A, and only three cooks in City B, came up to the median of the trade as a whole. In the blacksmith and horseshoeing trades, two trades with low educational attainments as well as low wages, there have been practically no apprentices learning the trade during the past several years.

Relation of the Grade At Leaving School to the Average Kind of Trade Characteristic of the Grade

From the medians of Table II, we are able to see that there is some relation between the grade at which a boy leaves school and the trade which he enters. We wish now to secure some measure of that relation.

In order to do this, the trades were arranged in order of their total median grades, and rank values were given to each trade as shown in the second column of Table II. Horseshoers, with the lowest total median is given the rank, 1; carburetor man, at the other end of the range, is given the rank of 30. Then, from the crude number distribution by grades and by trades was calculated the average trade rank of each grade.

The sampling is of course quite small in the very lowest and very highest grades, and could undoubtedly be improved by securing larger numbers at these grades. From the results of the above computation, the average trade entered by third grade men may be assumed to be about that below structural steel erecting, if we consider the number of cases large enough

GRADE III HAS 11 CASES AS FOLLOWS

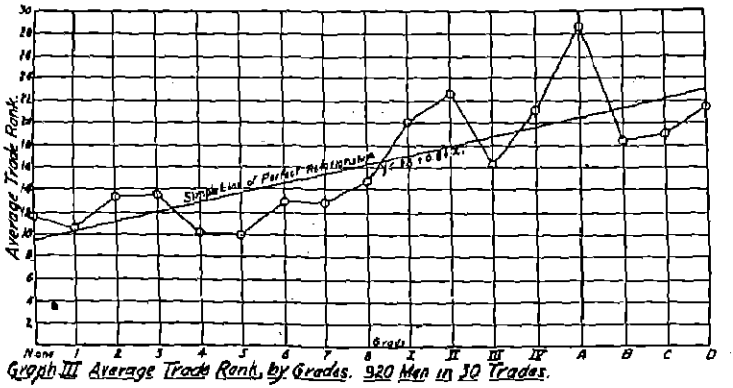
No. of Cases	Trade	Trade Rank	Product Cases x Rank
1	Cobbler	2	2.0
2	Carpenter	5.5	11.0
2	Switchboard Rep.	29	58.0
2	Telephone Lineman	5.5	11.0
1	Locomotive Repair	9.5	9.5
1	Elec. Inst. Repair	20.5	20.5
1	Inside Wireman	20.5	20.5
1	Powerhouse A. C.	14.5	14.5
11			147.0
		Average	13.4

to justify such a conclusion. The average trade by grade, the average deviation in trade rank from these averages and the number of cases upon which they are based are all shown in Table IV. We note a general tendency of these average trade ranks to increase with increase in the amount of education as we should expect. If the relationship were a perfect

TABLE IV. CORRELATION OF THE AVERAGE TRADE PER GRADE WITH THE GRADES

Grade	To Correlate with rank	Average Trade	Average Vi= Deviation	Number of Cases
None	0	11.6	8.7	9
1	1	10.5	6.7	3
2	2	13.3	5.2	6
3	3	13.4	8.4	11
4	4	10.1	5.0	26
5	5	10.0	5.0	43
6	6	12.9	6.6	108
7	7	12.8	7.0	91
8	8	14.7	6.6	417
I H. S.	9	20.1	5.7	56
II H. S.	10	22.6	6.4	59
III H. S.	11	16.2	6.6	25
IV H. S.	12	21.1	6.1	45
A College	13	28.6	1.4	5
B College	14	18.5	3.3	3
C College	15	19.0	2.2	4
D College	16	21.3	6.2	10
Average		16.3	5.7	Total
(Pearson) r =	.79 ± .06		...	920

one we should have, as the simplest (example), a straight line curve when these averages are plotted against the grades themselves. When, however, we plot these (Graph III) we find much deviation from a straight line, due perhaps in some cases to the smallness of the numbers and also to the fact that we should actually expect more deviation at the extremes of the series if the grade at leaving school is the result of more than one factor as it is known to be. Not only should these averages be such that each succeeding one should be larger than the preceding but they should be in some series order, the simplest of which is of course an arithmetical series, or 0, 1, 2, 3 — 16 order. One measure of this is a correlation ratio. The Pearson correlation, r, between this arithmetical



Graph III Average Trade Rank by Grades. 920 Men in 30 Trades.

series and the average trade values by grades is 0.79 with a probable error of 0.06.

By employing the method of *least squares*, we may give large amount of weight to the extreme deviations which seems desirable, and may determine a straight line equation which will best fit the defective data at hand.

The equation of the line of least square deviation is $y = 9.51 + 0.84\,x$. This line is plotted on Graph III.

Mentality and its Relation to the Trades and Trade Status.

We have previously pointed out the probable close relationship of mentality to education. In the present work we have often felt convinced that the only difference between many journeymen and experts was one of intelligence. This is often so considered by the foreman in charge of the men. A few

quotations of foremen are here given as examples of the layman's method of expressing the mental disabilities well known to the clinical psychologist by other terms:

"He makes good journeyman wages and works all right as long as the machine works all right; but if anything goes wrong, he can't fix it."

"He hasn't the head for an expert, although he is just as good a workman as this other man (an expert)."

"This man works all right so long as I tell him what to do."

"He isn't 'all there' when it comes to doing new jobs."

"The man can work all right as long as he is told what to do but he can't reason out things for himself."

"He can follow a drawing all right, but he can't figure out things."

If we examine certain of our medians we may get some interesting light on the practical nature of the differences between the different trade classes.

The median grade of journeymen plus experts is 8.38; that of journeymen plus experts plus apprentices is 8.39. This suggests that journeymen and experts are but later differentiations of the one class, apprentices or beginners, the better educated class falling in general into the expert group and the poorer educated class falling into the journeyman group. Furthermore the median of journeymen plus experts (8.38) is an exact average between journeymen and experts (8.24 to 8.51). *This suggest that journeymen may also be further* split up into two classes. For, if the average of the median of 309 J + median of 363 E = median of the distribution of 672 journeymen and experts combined as one distribution; then, since, only if there were a real basic difference in the education of journeymen and experts and equal numbers of each if the groups were normal groups, would this relationship hold true, a plausible explanation under the circumstances is that the journeymen are made up of an undue proportion of the poorer educated which may arbitrarily be considered as one class, and of a small proportion of better educated men which may be considered another class.

In the light of the foregoing opinions of foremen we would interpret this as indicating two classes of mentality to be distinguished among the journeymen; first, those of high mentality who are classed as journeymen merely because of having not yet had sufficient years experience to be classed as experts, and secondly those "doomed to be journeymen all their days."

If, then, we should exclude from the journeyman median those who are future experts *in embryo,* we should find a much lower median of education of the remaining class.

Greater average deviation by trades (Table II) from the median of the status would seem to indicate that the journeyman period is a period of adjustment of the workmen to jobs. The poorer educated and more poorly mentally endowed, after a few years of trial at work begun, we must think, on a more or less hit-and-miss principle of trade selection, are forced by the competition of better educated and better mentally endowed men to make new adjustments to their industrial life when time has given opportunity for their mental shortcomings to become apparent. At least one large industry of the country removes from the payroll any man who fails to progress beyond a given point in learning the trade and is not thus continually becoming more useful to the company. The mentally inferior are thus practically forced to accept the less desirable and less responsible positions.

The curves of Graph II, while not based upon a sufficient number of cases to be taken as positive evidence, seem possibly to indicate that the conflict between mentality and ability to progress well in school work comes later, on the average, in the case of apprentices and experts (about I or II H. S.) than in the case of journeymen. The peculiar rises in the elimination curves of the apprentices and experts at II H. S., a phenomenon not previously noticed and also absent in the case of the journeymen, while the ordinary 6 and 7 grade peculiarity is absent in both the expert and apprentice curves, would certainly indicate such a conclusion if the existence of these peculiarities in the curves should be confirmed by future investigation.

Unemployed men have several times been shown to have lower mentality than the population taken as a whole. Since these differ presumably from the whole population in education mainly in the greater proportion of poorly educated men we conclude that the unemployed group contains an undue proportion of the "doomed to be journeymen" class of men. Periodical calls at the employment office is thus but one of the incidents in the life of such unfortunate individuals that serve to indicate to the outside observer the readjustment of mentality to the demands of the trade. Of men who change their trade after a few years work we should expect the larger part to take up new trades of a similar character. Those who do so would be the class with the greatest amount of mental deficiency. We believe that the better endowed man is more apt

to remain in even a very simple trade when he discovers that he has entered a trade not requiring all of his mental ability, for the simple reason that on account of that superior ability he is able to make rapid progress in the trade and win promotion rapidly. This is a problem for future investigation.

Relation of the Education of Tradesmen to That of the Professions and of Laborers.

The best known professions are quite commonly recognized as requiring a college degree as a necessary prerequisite for entrance into the profession. The semi-professions would also undoubtedly rank high in educational attainments.

The education of laborers in its relation to that of tradesmen is not so clearly apparent from merely our knowledge of the arbitrary standards demanded, as for example, of anyone who would follow the professions. Laborers require surely the least of knowledge gained from school study. Do then the ranks of the laborer contain only men of a small amount of education, or do they contain as well many men of a fair or very large amount of education?

In an attempt to answer this question, we present here the tabulated results by grades of 50 each white and colored laborers chosen at random from the files of the Personnel office, Camp Dix, New Jersey.[4]

PERCENTAGE DISTRIBUTIONS BY GRADE AT LEAVING SCHOOL OF FIFTY EACH WHITE AND COLORED LABORERS AT CAMP DIX, NEW JERSEY

Color	Per Cent at Grade:— None	1	2	3	4	5	6	7	8	I	II	Total	Median Grade
White.....	56	14	6	14	6	2		2				100	0.91
Colored.....	44	16	8	10	2	6	6	4	2		2	100	1.44
Both.....	50	15	7	12	4	4	3	3	1		1	100	1.00

The results here presented may be too low for laborers in general, due to the fact that most of the white laborers were foreigners, and also to the fact of the great demand at this time for men of fair ability in the more skilled trades. These together account for the lower median grade of the white laborers. Native born white laborers may be expected to have a slightly better education perhaps than that here indicated for the colored men.

The data is quite conclusive evidence that the occupation of laborer, the one requiring least skill and ability, is held

[4] We gladly acknowledge our indebtedness to First Lieutenant E. S. Gregory, Jr. for his cooperation in securing this data.

generally by persons of very inferior education. Eighty-eight per cent of the laborers have less than a fifth grade education On the other hand, over eighty-nine per cent, or a larger percentage, of all the tradesmen are found with an education equal to or better than a fifth grade education.

Industrial Selection of Men and Trade Guidance by the Amount of Education Attained.

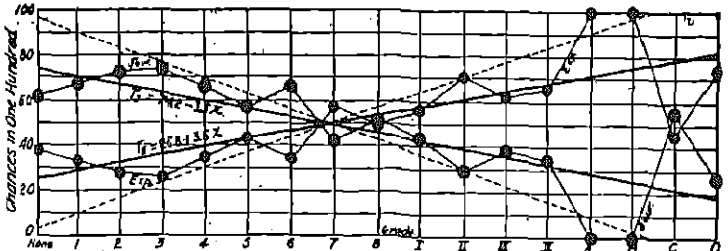
From the employer's point of view, of two men of equal skill and trade ability (irrespective of years of experience) the more valuable man would ordinarily be the more intelligent one. Given equal chance to learn the trade, this would ordinarily mean the man who could learn the trade more quickly. In some cases, however, there is a distinct demand for the moron type of individual. "He is better satisfied and will take more from his foreman." In such cases High School graduates would scarcely be in demand at all. A low degree of education would usually pick out such an individual; this however, is seldom needed as it is usually possible to pick out the extreme types of intelligence by simple methods much better than by the criterion of education.

In those trades demanding a high degree of skill and responsibility, high school men would seem to be desirable. Our figures show that in only 41 per cent of the trades considered are there more than one-fourth of the men in the trade who have even entered high school. Of the experts, in 55 per cent of the trades have one-fourth or more of the men entered high school. This clearly indicates where "expert timber" is grown in greatest abundance. Of the journeymen, only 27 per cent of the trades have one-fourth or more of their men who have entered high school.

Only 14.9 per cent of journeymen have some high school education, while 27.8 per cent, or nearly twice as many, experts have some high school education. The experts exceed the apprentices by only 3.6 per cent on this score, due no doubt to the fact that many of the better educated men reaching the expert class go even higher into the executive positions where they are no longer considered tradesmen in the ordinary sense of the term, a man toiling at a trade. These figures are such as to further confirm our belief that among the journeymen we find a smaller percentage of well educated men.

From the percentages of Table I, we can, by comparison of the journeyman and expert percentages, roughly predict what will be a boy's chances, on the average, of eventually becoming an expert or of remaining a journeyman if he leaves school

at a given grade. The percentages for the journeymen are probably too low in the lower grades for an exact determination of this as has already been pointed out. If we plot the percentages of Columns 8 and 9 of Table I as percentages of their sum, we secure the curves of Graph IV. The curve, Exp. represents the number of chances out of one hundred a boy has of becoming an expert if he leaves school at a particular grade. The curve, Jour., gives the same for his chances of becoming and remaining a journeyman. This data can also be subjected to the method of *least squares* whereupon we secure the straight line equations, T_{-E} and T_{-J}, (Graph IV) which are perhaps as representative of the facts, and are much simpler to comprehend. By reference to these curves, we may say that if a boy leaves school before the seventh grade, the chances



Graph IV.- A Boy's Chances of Becoming an Expert Dependent upon Education Attained

are greater that he will be a journeyman than that he will become an expert, the chances of permanently remaining a journeyman being greater the less education he has. After the seventh grade, where the chances are about even, the chances for eventually becoming an expert gradually increase until in about four out of every five cases will a college graduate become an expert. Were the percentages of Table I corrected so as to include in journeymen only those "doomed to be journeymen" we might expect the true position of the curves, T_{-J} and T_{-E}, to assume somewhat the positions of the dotted lines of Graph IV.

Trade Education of the Men Examined.

Few of these men had any original trade instruction. Such trade instruction as they received was largely at the hands of hard experience or of still more exacting and unappreciative foremen. The method of instruction is left almost wholly to the shops themselves. Such instruction must then necessarily be more or less of the same calibre as the men who work

at the trade and who instruct the incoming apprentice. The writers have become quite convinced that the time required to produce a journeyman workman—five years in many cases—is entirely too long with men of average intelligence. Not only could the average person learn the trade much quicker than this if given the opportunity, but in many cases he is prevented from so doing by a lock-step system of apprenticeship. The successful experiments now being undertaken by the War Department, wherein men of fair intelligence and general training are taught just such trades as these in short courses is well worth watching by our educational and industrial leaders.

We have the testimony of the educational director of one of the largest industrial concerns of the country to the effect that "a two year apprentice with organized school-shop instruction can not only produce as good work as a five or six year journeyman without the instruction but he is more valuable to the company because he knows "why he does things." Or as another very intelligent foreman expressed it, "the real expert is the man who can run a machine when the machine fails to work; anyone can run a machine when the machine is working properly. The expert knows the 'why' of the machine." Perhaps the first man overlooked the fact that many of the boys who come into his hands are the boys of better intelligence to begin with and would have done fairly well without the school-shop instruction, as this company is rather careful in choosing its employees. The second man may have under-estimated the value of the human element in the production with "things going well."

Inasmuch as most men differ greatly in their rates of learning the "why" rather than in their ability to eventually learn these comparatively simple trades, it seems quite evident that organized trade instruction in the shops or in the schools would do much to alter the proportion of the journeymen and experts with the favorable balance on the expert side.

Finally, if the results found for the education of laborers be taken to represent the education of laborers in general, and if those found for tradesmen be taken to represent the education of tradesmen in general, we can see the broader relations of education to the three divisions of occupations as laborers, tradesmen and professional men. A boy with no education at all, or even with as much as a third or fourth grade education is practically condemned to the life of a laborer. From a third or fourth grade education on up but with less than a seventh grade education the chances for his becoming a journeyman

tradesman only, are greater than his chances of becoming an expert tradesman. With better than a seventh grade education his chances are greatly in favor of eventually becoming an expert tradesman. And, finally, if he would be a professional man it is necessary that he acquire a college degree. That these statements are true only for the general type of man found in these divisions is clearly apparent. No one at all acquainted with the general type of laborer would maintain that by spending twenty years in an effort to educate him through college could one make a professional man out of him. He lacks the general intelligence of the professional man. And, on the other hand, not all men with college degrees make successful professional men, nor even always successful tradesmen; nevertheless a college degree is generally good industrial insurance.

PRINCIPLES UNDERLYING THE CLASSIFICATION OF MEN [1]

By TRUMAN L. KELLEY, Teachers College Columbia University.

PART I.

The opportunity to appear before this body suggested making a report upon some one of the phases of army work with which I have been connected, or of presenting certain conclusions and principles which have been developed or strengthened by my total War Department experience. I have chosen the latter course because I believe my report will be of more value if I present some of the principles of selection underlying war procedure which are equally applicable to peace conditions. My experience in connection with tests for illiterates, trade tests, and the selection of ground and flying school men, entirely supports the conclusions which will be reported, but I shall choose another field, that of selection of officer material from S. A. T. C. units, to illustrate the points made, because, in my judgment, the principles adopted in the selection and classification of S. A. T. C. men are in advance of those used in any other field of classification. The problem of classification in the Students Army Training Corps followed directly from its main purpose which was to provide for the securing of material fit for officer training.

It was anticipated that from thirty to forty thousand men a year would be selected from the S. A. T. C. units for transfer to training in officers' training camps. Not only was it necessary to select these men, but it was desirable to assign them for training to those branches for which they were best fitted and in which they were most needed.

Had proverbial methods for handling a case like this been followed each corps would have conducted a separate and intense recruiting campaign in the S. A. T. C. units—the corps first on the field or the one that held out the most alluring picture of the life of its officers getting the cream of the men, to the disadvantage of other corps and resulting in a net de-

[1] Part 1 was first presented in an addresss delivered at the Baltimore meeting of the American Psychological Association under the title "Principles Underlying the Classification of Men in the Students Army Training Corps." Part 2 is presented here for the first time.

crease in the efficiency of the army as a whole. It would also be expected that in lieu of objective methods personal bias of one sort or another would be largely operative in determining recommendations for officer training schools.

In addition to securing general officer material it was contemplated that quite a number of specialists, bacteriologists, chemists, ordnance experts, etc., would be drawn from the S. A. T. C.; also, that thirty or forty thousand men fitted for training in schools turning out non-commissioned officers would be found and that those remaining after these selections would be assigned as privates to regular line units.

This was the situation which Dr. Thorndike was called upon to solve. Accordingly the scheme which he drew up aimed to do the following things:

1. Pick specialists.
2. Divide the soldiers into three parts upon the basis of general merit as officer material, the upper part going to officers' training schools, the middle part to non-commissioned officers' schools, and the least meritorious part to camps, continuing upon the status of privates.
3. To do away with the necessity of each corps conducting a separate recruiting campaign.
4. To reduce to a minimum the tendency to exercise personal bias.
5. To be just to the soldiers, both from the standpoint of their abilities and from that of their interests.
6. To determine scores for each man selected for officer material indicative of his respective degrees of fitness for the different branches of the army.
7. To lead to an allotment of officer material to the different branches which both in number and quality would be appropriate to their needs.

The solution to so complex a problem demanded a procedure involving the closest synthesis between the abilities of men, the special needs of different branches of the service, and the numbers required in the branches.

The determination of whether a man is possessed of the special training necessary to fill a specialist job is not a difficult task, so that it was relatively easy to provide for the selection of these men. According to Dr. Thorndike's scheme, these were to be selected before the further classification was done.

The division of the remainder of the soldiers into three groups, and their classification for different branches, depends upon securing accurate ratings upon traits essential to general

fitness and distinctive of such capacities as are especially required in the different branches of the army. Dr. Thorndike's original plan, therefore, provided for rating a man in the following twelve qualities:

1. Command of men
2. Popularity
3. Military studies
4. Military practice
5. Intellect, including academic studies
6. Athletics
7. Athletic—mechanical ability
8. Business ability
9. System
10. Physique
11. Character
12. Scrupulousness

These qualities were chosen because (1) each one was essential in at least one of the branches of the service for which selection was to be made, (2) it was possible to secure fairly reliable ratings for each of them, and (3) in the main they measure quite discrete talents. Due to the withdrawal of certain branches of the service from those drawing from the S. A. T. C. the number of traits was later reduced to seven as follows:

1. Intellect, including academic studies
2. Character
3. Military studies and practice
4. Physique and athletics
5. Command of men
6. Athletic-mechanical ability
7. Scrupulousness

In order to secure ratings upon these traits provision was made for a Rating Board, and in order that the rating should be as reliable as possible these principles served as a guide in determining membership and procedure.

1. An average of a number of ratings is more reliable than a single rating and is less liable to be determined by personal bias.

2. An average of independent ratings is more significant than a concensus of opinion reached after discussion.

3. In order to be accurate, ratings must be made by men who are intimately acquainted with the daily lives of those rated.

4. In order to be amendable to interpretation, ratings when made by different men must be upon the same scale.

In order that men with the peculiar talents necessary for the different branches should be so assigned, a bidding for talents, i. e., for the twelve traits listed above, by the officers in charge of recruiting in the different branches, was secured. Each branch was entitled to one hundred units of talent and could divide them among the twelve traits listed in such proportion as it judged to be most advantageous to the particular needs of its corps.

Each man was given a score in each branch of the service, which was the sum of the products of his scores in the various traits and the bids for these traits as made by each corps. To these certain amounts were added on account of the man's individual preference, resulting in final fitness scores for each branch.

Finally an allotting plan provided in substance for the assignment of each man to the branch for which he was best fitted, unless the quota in such branch was filled, in which case he was assigned to the branch for which he was next best qualified.

This brief outline presents the problem which demanded solution and the principles which were quite evidently followed by Dr. Thorndike in meeting it. With this concrete plan before us we may turn to a consideration of such principles of selection as are universally sound.

Accurate classification depends upon the securing of such measures of fitness, before the men have been called upon to demonstrate their fitness by actual performance, as will correlate highly with the excellence of the later performance. The only method of proving that the classification is accurate is by finding a high degree of correlation between the diagnostic measure and the measure of performance.

But before there can be proof there must be something to prove and it is with a view to suggesting kinds of tests and measures that offer such promise as to make it worth while to test them out, that the factors making for high correlation will shortly be enumerated.

I would ask you to note that the method of proof is straightforward and yields conclusive evidence, and in order to keep the issues of this paper perfectly clear it is hereby disclaimed that any of the principles presented for judging of the appropriateness of diagnostic measures provides a second method of proof. They constitute suggestions as a guide in drawing

up tests in the first instance and criteria to use in estimating the excellence of measures where *proof* is impossible.

The facts of partial correlation show us that when measures of any sort are used together, i.e., their results summed, to obtain a single score diagnostic of some ability, three conditions as to the relationships of the separate measures determine their excellence. These conditions determine the magnitude of the partial correlations between the separate tests and performance, and are, other things being equal, (1) the higher the reliability coefficient of the test or measure the better, (2) the higher the correlation between the measure and performance the better, and (3) the lower the correlations with the other measures with which it is to be combined the better. More briefly; high reliability, high correlation with demonstrated efficiency, and low intercorrelations are the characteristics of a good diagnostic measure. This can be made clearer by an illustration.

Suppose the capacity to be measured is fitness for training as an infantry officer. Suppose scores in the following have been suggested as constituting a good set of diagnostic measures: (1) school marks; (2) teachers' estimates of general intelligence; (3) drill masters' estimates of courage; (4) drill masters' estimates of ability to command men; (5) grades received in military drill; (6) teachers estimates of character; and, (7) estimates made by students themselves of the degree of their interest in performing the duties of an infantry officer. The significance of each of these measures can be determined if records showing efficiency in a camp for infantry officers are available, in the case of men for whom the above scores were determined. The existent conditions of reliability, correlation, and intercorrelation determine this significance for each of the tests. For example, if the estimates of general intelligence are (a) fairly reliable, (b) correlate quite highly with demonstrated efficiency and (c) correlate still more highly with school marks and estimates of character and quite highly with the remaining four traits, they will turn out to be fair or poor, but not good measures for the purpose in hand. It is the high intercorrelations which prevent their being excellent measures.

These well known facts of partial correlation have been dwelt upon because they furnish the key for judging of the value of a test when the measures of demonstrated fitness are not available. It is probable that it is always possible to make more accurate estimates of the reliability of a test, its correlation with the ultimate criterion, and its intercorrelation, than

it is to estimate directly its value as a diagnostic measure. Thus if it is impossible to secure demonstrated fitness measures, and therefore the only method possible is that of estimating significance, a very much more reliable estimate can be made if it is based upon three estimates—one of reliability, one of correlation with the criterion, and one of independence—each of which is more or less within the experience of one familiar with test and mental relationships, than if made independently.

This guessing method however can usually be improved upon, for generally actual coefficients of reliability and coefficients giving the intercorrelations are easily secured, even when it is out of the question to secure measures of correlation with real efficiency. In this case estimaes as to the latter only, need to be made.

At times a knowledge of reliability and intercorrelation may solve the whole problem so far as the single test is concerned. If, in the example cited, the coefficient of reliability for estimates of courage should turn out to be .10 or less—and the determination of this coefficient is a very simple matter—these measures can be immediately discarded as being practically worthless; and this in spite of the fact that courage itself may be the most important trait for an infantry officer to possess. Again, if the correlation between drill masters' estimates of ability to command men and grades received in military drill should be .95 or above, it would be immediately apparent that nothing was to be gained by keeping both measures and one could be discarded.

Without stopping to consider each of the seven traits, it may be pointed out that probably the estimates of interest made by the students themselves stand high as judged by each of the three criteria and accordingly such estimates may be expected to be highly important in diagnosing fitness for officer training.

The problem cited, tho neither more nor less difficult than the ordinary vocational selection problem, shows the difficulties in approximating, without knowledge of the partial correlations, the importance of a measure. Too many relationships are involved to permit of a really accurate evaluation by means of logical analysis alone. But, as mentioned, the *attempt* to estimate and interpret *all* the relationships relevant to partial correlation is particularly serviceable in *suggesting* measures which may be expected to prove valuable, since of course a method involving proof is not available at this stage of the procedure.

To this point I have considered the means of determining the excellence of measures of individuals as evidence of fit-

ness for some one particular job. I shall now consider the excellence of measures of success in jobs as evidence of the possession by individuals of certain traits. Before, the question was to judge of the probable degree of success in a job of a man with certain known, i. e., measured traits; now the question is to determine the probable degree in which a trait is possessed by a man who has known measures of job successes to his credit. This question is not merely an academic one, for it must be solved in some way wherever a certain number of men are to be allotted to a number of different jobs. The principles of partial correlation are operative here as before. Just as a number of men with various standings in several traits show different degrees of efficiency for one job, so *one* man will show different degrees of efficiency in a *number* of jobs. The method of determining which of the several jobs he will best fit is in theory just the same as the preceding, interchanging the variables. In actual working out it would not be as simple as implied but the details need not here concern us.

If the two processes are welded into one and made a single analytical function we have the solution to just such a problem as that of a firm having 100 vacancies in different departments to be filled by 100 men, each man being sufficiently meritorious to warrant some position.

If the vacancies are all of equal importance the best solution would be that which would result in the minimum number (or more accurately amount) of displacements of the 100 men—displacements to be measured in some such way as the number of transfers from one position to another until a position is found which is filled in an excellent manner. As positions ordinarily are not of equal importance, but roughly of an importance represented by the salary paid, a close approximation to the best placement of the men would be that which would result in the minimum number of dollar displacements—a dollar displacement being obtained from individual displacements by multiplying each of the latter by the respective salaries involved. It is obvious that some such weighting of individual misplacements is needed for it is much more serious for a $25 a day executive to be in a position that he is not fitted for than for a $1 a day messenger to be so placed. The problem can be readily expressed in terms of the method of least squares, or minimum errors, tho its solution statistically is complicated.

Through the genius of Dr. Thorndike a solution to this very problem evidently based upon the principles of partial corre-

lation where possible and empirical where necessary, has been most successfully attempted in his scheme just described for the selection and classification of S. A. T. C. men. His problem was not quite as general as above depicted, for in a practical sense it may be assumed that all officerships of the same rank are of equal importance, i. e. it is just as serious from the standpoint of army efficiency to have an inefficient lieutenant in the infantry as in the signal corps. The Committee on Education and Special Training of the War Department has a few copies of the scheme, as it was revised by a Committee of Supervisors of Personnel Methods and as altered in certain details by the administrative officers of the Committee on Education and Special Training, which can be obtained by those interested. In the form presented it is probably slightly less sound theoretically than as left by Dr. Thorndike, tho simpler in administration. The capitulation of Germany prevented this scheme from being put into general operation. It was however tried out in two institutions with very satisfactory results.

To summarize the salient features—it involves:

1. A method of elimination from consideration of such men as are not fit for any officer post.
2. A rating of men with respect to possession of certain traits—traits which promise to be important when judged by the criteria previously mentioned.
3. A rating of officer jobs in the different branches of the army with respect to the needs of the branches for men possessing the given traits.
4. A scheme for utilizing these ratings in allotment whereby the men would be best placed and the jobs best filled considering the total supply of available officer material. The aim was to make misplacements a minimum, in fact that there should be no misplacement whatever except as inaccuracies in the original ratings just mentioned should lead to it.

Items (3) and (4) are original contributions to the problem of classification and as I interpret them, are in harmony with well established statistical principles for a fitting of talent to need.

The possibility of utilizing such a scheme for peace purposes in industrial and educational classification is far-reaching, for every man is better fitted for some one place than any other, tho of course this will occasionally be no more honorable than a place in an institution for the feeble-minded. The

discovery of this 'best place' for an individual is an obligation which is definitely thrust upon the psychologist to solve.

Part II. The Analysis of Capacity.

It has been shown in the preceding part that a proper fitting of talent to need is only to be accomplished when the job analysis and the individual analysis show that there are fitting degrees of talents in the individual to meet established needs in the job.

This part will concern itself with the lines along which job analysis on the one hand and individual analysis upon the other should proceed. The categories into which the elements involved in either analysis may be divided, are not so distinct as to be invariable, but certain divisions will prove of greater utility than others. What these are in detail depends upon the task for which selection is to be made. There are, however, two fields for which men are selected which are so different as to make it necessary to consider them separately. They are selection upon the basis of acquired habits, such for example as on the ground of trade skill, and selection upon the basis of *ability* to learn, such as selection for special training in trade or other schools. Occasionally selection involves both of these bases, but as the criteria for judging of the fitness of tests or methods are different in the two cases, and as commonly but one basis is marked in any given problem of selection, the two will be considered separately.

As an illustration of the relative independence of the two types of selection in practical work, it may be noted that the Army Literacy Test and the Army Trade Tests select upon the basis of attainment, while the Army Intelligence Tests, the Scott plan for rating officers, the Thorndike plan for the selection of ground and flying school men, and the Thorndike plan for the selection of S. A. T. C. officer material, endeavor primarily to differentiate upon the basis of ability to learn.

Before considering these two kinds of selection the value of attempting a task analysis at all needs to be supported, for if efficiency in a large number of common tasks is altogether measured by the degree in which some single function, let us call it general intelligence, is present, and if general intelligence is a single thing a measure of which can be approximated in one or more ways, then it is quite superfluous to analyze the task, for such analysis could only result in revealing different ways in which the single capacity operates. Without becoming involved in theoretic considerations, it may be noted that there is a pragmatic acceptance of the single mental function theory wherever there is an acceptance of an intelli-

gence test, such as the Binet test or the Army Intelligence Test, as a means of differentiating between abilities practically irrespective of the purposes to be served. Whether a general intelligence measure is all sufficient is a question which does not need to be left to opinion, for it is merely a statistical problem to determine the degree of correlation between a task and measures which very evidently are not intelligence measures; or, if in part measures of intelligence, to determine the partial correlations involved.

Measures other than of general intelligence have been found to correlate with success in a very large number of lines. That it may be expected that such measures exist for all the important activities in life is supported by the very partial degree to which general intelligence determines the vocation which one selects. To illustrate: In a pamphlet entitled Army Mental Tests, distributions in Army Intelligence Test scores for men following different vocations are given by the Division of Psychology of the Surgeon General's Office. These data are presented in the form of a chart giving for each vocation the lower quartile, median, and upper quartile scores. The median score of 'laborers' is lowest and of 'engineer officers' highest. From this chart it is possible to closely estimate the spread in the scores for each group and for all the groups combined. If the standard deviation of the scores for all groups combined, weighted as frequencies in civil life warrant, is taken as 1.00 it is found by scaling and simple calculation that the standard deviations for the separate vocations do not vary greatly and are approximately ⅞. This large standard deviation for the separate vocations, taken in conjunction with the fact that the Army Intelligence Test has a reliability coefficient of about .97, proves that factors other than general intelligence are very largely operative in determining selection of a vocation. (In this discussion 'general intelligence' is used synonymously with 'Army Intelligence Test score.') The mathematical proof of this statement is accomplished by securing measures showing the relative importance of general intelligence and other factors entering into selection of a vocation. [2]

[2] As the proof of this statement depends upon certain facts of partial correlation, one not acquainted with them may skip the mathematical analysis, simply noting the values of the final coefficients of correlation obtained and given at the end of the paragraph. The following formulae, given in Yule, G. Udny—*Introduction to the Theory of Statistics;* Kelley, Truman L.—*Educational Guidance;* or Kelley, Truman L.—*Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations* (out of print); are utilized:

It may help in following the illustration to think of the vocations as ranking in the accompanying order, which is that given by the Army Intelligence Test scores, altho the true ranking would be on the basis of the sum total of abilities, interests, wages, geographical positions, etc., etc., which determine participation in a vocation:

Laborers
Tailors
Cobblers
General miners

Painters
General blacksmiths
Bridge carpenters
:
Telegraphers
R. R. clerks
Filing clerks
:
Medical officers
Army chaplains
Engineer officers

Let us designate by a variable, x_1, position of any vocation in such a true ranking. Let us designate by x_2 intelligence test position and by x_3 position in a composite of *all* the *other* factors which determine an individual's selection of a vocation. x_1 is related to x_2 and x_3, using the usual regression equation notation as follows:

1. The regression equation involving three variables:

$$x_1 = b_{12 \cdot 3}\, x_2 + b_{13 \cdot 2}\, x_3$$

2. The standard deviation of the errors of estimate in the case of three variables:

$$\sigma_{1 \cdot 23} = \sigma_1 \sqrt{1 - r^2_{12}} \sqrt{1 - r^2_{13 \cdot 2}}$$

3. The value of a partial coefficient of correlation:

$$r_{12 \cdot 3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r^2_{13}}\ \sqrt{1 - r^2_{23}}}$$

4. The standard deviation of the errors of estimate (or of the arrays) in the case of two variables:

$$\sigma_{1 \cdot 2} = \sigma_1 \sqrt{1 - r^2_{12}}$$

$$x_1 = b_{12.3}\ x_2 + b_{13.2}\ x_3$$

and since x_3 comprises *all* the factors other than general intelligence which affect vocational choice, x_1 is completely determined when x_2 and x_3 are known, i. e., $\sigma_{1.23}$, the standard deviation of the errors of estimate, would in that case equal 0. However,

$$\sigma_{1.23} = \sigma_1 \sqrt{1-r^2_{12}}\ \sqrt{1-r^2_{13.2}}$$

and this can only equal zero in case $r_{13.2}=1$. Accordingly,

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r^2_{12}}\ \sqrt{1-r^2_{23}}} = 1 \quad \text{Therefore,}$$

$$r_{13} - r_{12}r_{23} = \sqrt{1-r^2_{12}}\ \sqrt{1-r^2_{23}}$$

Since x_3 comprises factors *other* than x_2, x_2 and x_3 are completely independent, or $r_{23}=0$, and the last equation becomes,

$$r_{13} = \sqrt{1-r^2_{12}}$$

r_{12} may be determined from the standard deviations reported above, since

$\sigma_{2.1} = \sigma_2 \sqrt{1-r^2_{12}}$, in which $\sigma_{2.1} = \frac{4}{5}$ and $\sigma_2 = 1.00$.

Solving, $r_{12} = .484$, the correlation between general intelligence and vocational choice. Continuing,

$$r_{13} = \frac{\sigma_{2.1}}{\sigma_2} = .875^{3}$$

This is the correlation between the 'other factors' and vocational choice. As its derivation has shown, it is essentially a function of x_1 and x_2, since x_3 has only been introduced for the sake of the argument and is not a known factor, and just as .484 is the coefficient of correlation between intelligence and vocational choice, so may .875 be called the coefficient of alienation between the same two things. If we designate this coefficient by k_{12} the simple relationship, $k^2_{12} + r^2_{12} = 1$, always holds.

In view of this high correlation between these other factors and vocational choice it is eminently worth while to attempt to discover what they are, and not rest content with such light as measures of general intelligence alone give for vocational or educational guidance. A task analysis should throw light upon what constitutes this very large field which is not general intelligence but which correlates with vocational selection to the extent of .875.

[3] If the reliability of the Army Intelligence test, .97, were taken into account this figure would be a trifle lower.

For the successful pursuit of a trade (a) certain habits are demanded, (b) certain concrete items of information, (c) a certain degree of initiative and originality—typically represented when a worker is called upon to decide which of his trade habits is appropriate in a given new situation, and (d) in addition to these elements of ability there must be a certain willingness to perform the duties of the trade—in other words certain increments of satisfaction must result from their performance in order to insure perseverance and contentment.

(a) The measurement of habits is readily accomplished by means of performance tests involving the habits in question. In a trade test it is not feasible to actually measure each of the habits exercised by one proficient in a trade, so a sampling of the habits is resorted to. Certain samplings are much better than others which require the same amount of time. That there may be a sharp differentiation between trade and trade, the particular habits sampled should be as typical of the specific trade as possible; as largely dependent upon habit and as little dependent upon general common sense as possible; as closely allied to habits not measured but involved in the trade as possible; while at the same time they should be as unlike the other habits which are measured as possible. Meeting these conditions will insure that the field of the trade is covered and not simply that some small portion is measured again and again, and that the test does not measure concrete items of information or initiative, each of which can be much more expeditiously and accurately measured by verbal or written tests. Of course such considerations as material and time available for testing, and ease and accuracy of scoring also enter in. If the time available for testing is short satisfactory results can not be expected from a performance test, i. e., a test aiming to ascertain the efficiency of trade habits. A verbal or written stimulus is adequate to call forth specific items of knowledge, but a habit is ordinarily set off by a situation brought about by the worker himself in the performance of a routine. It takes skill and much time to involve the worker in such situations as will betray the trade habits which it is desired to examine. This is particularly true if it is desired to measure a number of habits in a single test. As a consequence it is not to be expected that a short trade test of an hour's duration will correlate as highly with efficiency as a fifteen minute oral or written test measuring concrete information; and in fact this situation has been repeatedly reported as existing in the case of different army trade tests. However this changes rather than lessens the importance of the performance

test, for, if carefully devised a second hour in a performance test can be made to yield almost as much new relevant information about a man's trade skill as the first hour; whereas a second fifteen minutes in an oral or written test gives a very decidedly decreased return from that of the first fifteen minutes. In all probability, given excellent tests in both cases, some such elation as this holds:

A 15 minute oral test gives more information as to trade skill than a 15 minute performance test.
A 1 hour oral test gives more information as to trade skill than a 1 hour performance test.
A 2 hour oral test gives equal information as to trade skill to that of a 2 hour performance test.
A 4 hour oral test gives less information as to trade skill than a 4 hour performance test.

That the criteria for judging of the excellence of a trade test enumerated in the preceding paragraph are not generally accepted by people actually making selections, was forcibly brought to the attention of the writer by a civil service examiner who said that he could tell by the way in which an acetylene welder picked up his torch, his degree of skill. The writer, as a witness to just such an examination, listened attentively to the examiner's definite and assured description of the workman's ability, and being both a guest and a novice in welding, expressed great interest in the examination though mentally noting that several trades involve handling torches; that the picking up of a torch is hardly a matter of habit, as the location of the work with respect to the torch varies with the job; that the degree of merit shown in the picking up process is entirely a matter of the subjective opinion of the examiner; etc., etc. However, to date the case is in favor of the examiner, for neither he nor the writer has ever been told that the man did not have the exact degree of skill attributed to him. The man, the examiner, and the writer have, like gentlemen, refrained from concerning themselves with each other's business.

This story is typical and not exceptional, for every examiner, be he a trade tester or a college professor, who does not check up on his classification by studying the later performance of those clasified, is confirmed in his procedure though it contain essentially and easily ascertainable erroneous devices. The less the test, in its giving and scoring, (but not in its derivation) depends upon the trade expertness of the examiner, generally the better—certainly the better from the standpoint of objectivity and uniformity or reliability of measurement.

(b) The measurement of specific items of information which are pertinent to a trade can be accomplished by means of an oral or written questionnaire, or by other means such as appropriate picture completion tests. Certain important difficulties may be mentioned: One is due to the fact that trade terms are very fluent—the word "plug" probably means a score of specific things in as many trades, and probably also no single one of these twenty "plugs" is universably known by that name. Another difficulty is due to the fact that foreigners have a different nomenclature—one which is not a literal translation of the English. Another is due to the difference in procedure in different parts of the country—carpenters trained in New York, Georgia, and California will answer the question "What is done first in building a house" in very different ways and each in a way which is right according to his training. Another is due to the ease with which people can be coached upon items of specific information—an individual with the barest trace of literacy can be taught to correctly answer the following written questions: "What is your name?" "In what country were you born?" "In what State in the Union is your domicile?" "Has your abode been restricted to the confines of the aforementioned habitat for a decennium?" Coaching is nearly as easy in other specific information tests. Questionnaires based upon pictures and involving few words (extensively used in the Army Trade Tests), picture completion tests, and picture matching tests seem to be as little subject to the diffculties mentioned as any.

(c) The measurement of such initiative and originality as is needed in a trade may be approached in a number of ways, all more or less similar to the usual means of measuring general intelligence. A specific variation from this which promises to yield exceptional return, is that of making trade material and procedure the content of the test rather than the usual general intelligence material which is considered to be common to the experience of Americans. Reasoning tests involving engineering material were used by Thorndike in his study of Engineering ability, (conducted for the Carnegie Foundation for the Advancement of Teaching, report of which is published by the Foundation) with very satisfactory results.

(d) There is little reason to believe that the measurement of interest should prove a serious obstacle in the case of trade testing, for it should be possible to ascertain by means of a questionnaire the degree of satisfaction which in the past has resulted from the practice of the trade. The amount of satisfaction found in the work in the past and the opportunity the

trade gives for realizing present aspirations provide a prognosis for the future.

Turning from the measurement of an acquired facility to react, to the measurement of facility to learn to react, a radically different situation is presented. The measurement of concrete items of information is of no value in itself, and only becomes diagnostic of ability to learn in so far as it reveals a necesary foundation for the learning process, and in so far as the learning of concrete items in the past correlates with the ability to learn new concrete items in the future. The measurement of habits holds a similar position. Abilities (a) and (b) are of less, and abilities (c) and (d) of greater significance, than before, but the measurement of ability to learn is so different from the measurement of attainment that an entirely different division into elements is called for.

The measurement of ability to learn may properly be divided into: (a) The measurement of prerequisites, ground work, or necessary foundations upon which to build; (b) The measurement of accomplishment in itself not essential but which has been attained as a result of an activity which correlates highly with the activity necessary in learning the task in hand, e. g. if there is a high degree of correlation between the ease with which an American can learn French and the ease with which he can learn Chinese, a measurement of the time spent in learning French and facility acquired would give evidence as to the probable success in learning Chinese; (c) the measurement of ability to learn the new task by tests which are as nearly as possible samples of operations demanded in the task itself; and (d) the measurement of interest.

(a) The utility of a measurement of prerequisites must be considered in connection with the question of transfer of training. It has been thoroughly demonstrated in the training of mechanicians and engineering officers in the army that a great many of the foundations assumed in peace times to be necessary for proper training are unessential. The writer has seen commissioned and non-commissioned officers taught the theory of the internal mechanism and of the explosive charge of the hand grenade, on the assumption that these were necessary to the proper manipulation of the grenade in warfare. Other army experience has shown the futility of this. Though the case cited is more extreme than ordinarily found, it is to be expected that the same sort of absurdity maintains in every field in which extreme press of circumstances has not led to the ascertainment of what constitutes the real essentials. This condition has a parallel in the measurement-of-fitness field,

for there is a very large amount of examining which, in so far as it results in reliable measures, yields measures of traits which are not essential to the task examined for.

The question as to the essential or non-essential nature of a foundation is to be answered by partial correlation, i. e. by calculating the extent to which excellence in this ground work, *other things being equal,* is evidence of fitness in the task. Simple correlation of excellence in the ground work with success in the task is not evidence that the foundation chosen is a true prerequisite for the task, or even that knowledge of the foundation is evidence of probable fitness in the task, if other measures of fitness are to be used at the same time. In spite of these objections, the difficulty of securing under ordinary test conditions true samplings of a new task, and the relative ease of securing measures of success in past attainment, warrant the expectation that this method, if very judiciously used, will yield valuable information.

(b) The precautions consequent to the limited degree in which training transfers, mentioned above in connection with utilizing knowledge of prerequisites, are equally necessary in utilizing measures of attainment in tasks not a part of the task under investigation, but supposed to be dependent upon the same sort of mental capacity as is this task in question.

(c) There is no doubt that the ability to expeditiously learn to do a sample of a task, correlates very highly with the ability to learn to do a lot of such samples—i. e. the task itself, should it be entirely composed of such samples. The individual who can quickly memorize the connection between "house" and "domus" will be adept in memorizing one thousand such. Surely no prognosis of ability can be better founded than that based upon performance in a thorough sampling of the task; but ability alone does not determine fitness, for fitness is indissolubly connected with interest.

(d) An individual who takes interest in a test involving memorizing ten foreign words and stands high in it may be very averse to memorizing five thousand such and even still more averse to regularly practicing such an ability in translating the literature of a foreign language. Interest in learning under test conditions; interest in learning under school or apprenticeship conditions; and interest in the repeated and routine exercise of the association formed under ordinary vocational conditions, are distinct things which cannot be expected to correlate to a high degree. As an actual illustration, the case may be cited of a young man who in three months learned a trade, ordinarily accomplished thru a four years' apprentice-

ship. He practiced the trade as a journeyman for two months and then left to take up work giving opportunity for training in a profession rather less remunerative than the trade, showing that in spite of interest and ability in the learning process, there was little interest in the routine exercise of the thing learned.

It may be roughly true that the greater the ability to learn the greater the interest in learning, but this does not imply the greater the interest in repetition after learning. Initiative and originality lead to satisfaction in the learning process, but to annoyance in the routine repetitive process. From the standpoint of interest there must be a happy degree of initiative and originality for each vocation—just enough to lead to such satisfaction in the learning of the vocation as to carry the individual through the difficulties of learning but not so much as to make the annoyance of routine performance unbearable.

The preceding paragraph has considered interest as related to ability, for initiative and originality are largely intellectual factors capable of measurement. There are also interests definitely related to other abilities, such as interest in music, painting, etc., etc. Interests other than those dependent upon ability presumably exist and of course influence readiness to learn and to follow a given line of work. Their direct measurement seems almost impossible but in so far as they are constant and will operate in the future as they have in the past, their influence can be approximated by ascertaining the degrees of satisfaction which have accompanied learning and routine occupations of various kinds in the past.

In recalling the traits needed for the successful performance of a trade and those needed for the successful learning of a new task it will be noted that general intelligence is only approximately synonymous with initiative and originality in selecting appropriate trade habits, and with ability to learn new tasks. The other factors may be expected to be in part, possibly in large part, those which correlate so highly (.875), with vocational selection.

THE PRACTICAL "EFFICIENCY" OF A GROUP SCALE OF INTELLIGENCE

STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

By S. L. Pressey and L. W. Pressey

I. *The Standardization of Tests of Intelligence in Terms of Their Uses.*

In the September number of the *Journal of Applied Psychology* the writers described a group scale of intelligence for the measurement of the native endowment of school children. The scale has thus far been employed for two purposes (1) to aid in a census of the feeble-minded children in the schools of a certain Indiana county—this as part of a "mental survey" to determine the number of children needing institutional care in the county—and (2) as an aid in sorting children, according to ability, in the public schools. The present paper aims to present some facts relative to the efficiency of the scale as a whole, and of its separate parts, for these two purposes.

It should be pointed out at once that such a method of evaluating tests of general intelligence—obvious as it is—is somewhat unusual. The Binet and Point Scales have made us familiar with methods of estimating, more or less directly, mental ability in terms of amounts of development. The scales have been standardized by determining the average performance of children of different ages. And the value of a given test has commonly been estimated by the extent to which it distinguishes the average child of one age from the average child of another age.

Tests may also be evaluated by correlation with the results from a large group of other tests of high differential value. Standardization in terms of mental development is, if we may trust recent experimentation,[1] not wholly relevant to the problem. Tests which are differential of chronological age

[1] Brigham, Carl C. Two Studies in Mental Tests. *Psychol. Monographs*, Vol. 24, 1917, No. 1, p. 245. Doll, E. A. A Brief Binet Simon Scale. *Psychol. Clinic*, Vol. 11, 1917, pp. 197-211 and 254-62; also Pressey, S. L. and Cole, L. W. Are the Present Psychological Scales Reliable for the Use with Adults? *J. of Abn. Psychol.*, Dec. 1918.

are not necessarily the tests which are differential of degrees of mental ability at a given age. And evaluation of a test, by correlation with the results from a group of other tests, results in an "inbreeding" which may have the advantages, but certainly has also the dangers, incident to in breeding in the biological field. Doubtless a more refined mathematical handling may avoid such a danger, and give us our final method for evaluating tests of "general intelligence." But meanwhile it is interesting to see whether it may not be possible, by simple and more direct methods, to obtain some indication of the efficiency with which such tests serve two of the many purposes for which such scales are ordinarily employed.

The present paper will take up in order three questions. First, what is the efficiency of the group scale used, as a whole, in separating out from a total school population (a) feeble-minded of institutional grade, and (b) those children who are sufficiently subnormal or brilliant to deserve special educational treatment? Secondly, what is the contribution of each test in the scale to these differentiations? Finally, what is the efficiency of certain of the most differential of these tests, when grouped into a brief scale, for the purposes above mentioned, as compared with the whole scale?

The scale was first used as a means of first selection of cases, in a mental survey made to determine the number of mental defectives, needing institutional care, in a certain Indiana county. In order to find out what ratings, on the scale, ought to be considered suggestive of such mental defect, the simple expedient was adopted of surveying with the scale, the state school for the feeble-minded. The scale was, however, planned chiefly as an aid in discovering those school children who, because of exceptional dullness or exceptional brilliancy deserve special educational treatment. In order to determine how well the scale performed this service a special class for gifted children and a special class for subnormal children in Louisville, Kentucky, were examined, and retarded and accelerated children in the Bloomington, Indiana schools were compared—these two groups being taken to best represent, in a system not having special classes, the school's judgment as to those capable of superior work or not capable of doing the work of the average child.[2] A rough measure has

[2] 6-8 was taken as normal for the first grade, and so on up. With this standard, retardation and acceleration are phenomena sufficiently abnormal to have a definite significance as indicative of a child's mental capacity. Age at time of examination was used, this being considered satisfactory for the purpose in hand.

been obtained of the completeness with which the scale sets these special groups apart from the group of average children.

The study is based on results from 986 Bloomington, Indiana, school children, 200 cases at the state "School for Feeble-minded Youth," 23 children in a class for gifted children, and 25 in a class for subnormal children at Louisville. Data from two other cities surveyed (a total of nearly 2,800 cases) will also be referred to.[3]

II. *The Differential Value of the Scale as a Whole.*

Let us first examine the efficiency of the scale as a means of first selection, in the public schools, of cases sufficiently defective to need institutional care. The relation of the scores obtained at the state "School for Feeble-minded Youth" to the total distributions of scores for the Bloomington children may be indicated very briefly.

All cases in the institution, with the exception of a few adult "boys" living on a branch farm, who were sufficiently literate to take the examination (that is, all cases who had advanced beyond the second grade) were tested.[4] Of the total 416 cases in the institution 11-18 years old, only three made a rating above the lowest ten percentile for the school children of the same age. One of these three has a Stanford-Binet I. Q. of 78, for the others no ratings are available. One is eleven, the other two are brothers aged twelve. All three are orphan asylum children; the best children come, they say at the school, from the orphanages. There are only five cases rating above the five percentile for the school children. Of the thousand odd cases over eighteen, again three rate above the ten percentile for sixteen. One of these has an I. Q. of 81, another of 85. All three are delinquent girls.

Such results seem to be evidence of some definiteness that the scale measures general intelligence, if we define general intelligence as an intellectual quality lack of which, to any considerable degree, requires that the individual concerned be placed in an institution, and possession of which enables the

[3] The writers wish to acknowledge their obligation to Dr. George W. Bliss, Superintendent of the Fort Wayne "School for Feebleminded Youth," to Miss Henrietta V. Race, school psychologist in Louisville, and to the school officials and teachers of the systems studied, for the opportunity to obtain these data, and for their help and interest in the work.

[4] It should be added that all doubtful cases were also given the examination, and that some fifteen papers were thrown out, the examinations showing little that could be considered other than chance successes.

individual to continue in the environment. The scale distinguishes quite definitely these two groups. It should be said that the particular institution surveyed possesses a smaller number of young high-grade defectives than certain eastern institutions with which the writers are familiar. It may very well be that a survey of one of these schools would show the scale much less clear-cut in its differentiations than it appears at present. But the scale was not being used to help in an enumeration of cases which were to be put in an institution in New York or Massachusetts. It was being used to determine how many cases, in the schools of a particular county in Indiana, were mental defectives of institution type, as that type was defined in this particular state. The statement that a certain Bloomington school girl sixteen years old scored lower than 33 of the 69 girls of her age already in the institution and that she made a total score of 15 where cases in the institution ran up scores as high as 110, surely goes straight to the point in terms no one can misunderstand. [5] It should be added that this particular survey was part of a larger study undertaken by the State Board of Charities in an effort to demonstrate the pressing need of further provision for the care of mental defectives. To carry conviction to people not familiar with the significance of findings on the Binet Scale, the statement just given is more effective than any statement of I. Q. or even of mental age might be. It surely means more to the average individual than to say merely that this girl makes the lowest score of any school child of her age in her town. It means more than the statement that she grades lowest of 268 unselected school children in the three towns surveyed. The direct comparison with the group at Fort Wayne has a most immediate practical significance which cannot be escaped.

A special class for gifted children at Louisville, Ky., had at the time the writers visited it, 23 members, eight to ten years of age. Their I. Q.s on the Stanford Binet scale ranged from 1.23 to 1.69. But more important for the purpose of evaluation of the scale, in the immediately practical ways which have been attempted, is the fact that all these children had been for a year or more in this special class, doing roughly two years work in one, and had proven their capacity for such work. These children make on the group-scale scores which fall in the highest ten percentile for Bloomington, with

[5] This girl has a mental age of six, an I. Q. of 42. She is in the fifth grade because only in this room are the seats large enough for her.

three exceptions; these three are only slightly below the ten percentile. They are the three who have the lowest mental ages.[6]

In the special class for subnormal children at Louisville three out of twenty-five make scores above the ten percentile for their age. As the class contained moral delinquents as well as merely dull children, this is not surprising.

It would appear then, from these Louisville data, that the scale was of some efficiency in separating children who have already been especially diagnosed by the school as subnormal or brilliant, from a group of average children. The differentiation is not as clear-cut as might be wished. But, taking the groups as wholes, it would seem fair to infer that the scale does primarily measure that sort of ability, possession of which, to a marked degree, makes a child capable of exceptional school work, and lack of which necessitates special educational treatment. The brilliant class averages at the 98 percentile for the total Bloomington group; the subnormal group at the 4 percentile.

It seemed probable that there would be similar groups of dull and gifted children among the Bloomington school children. But since there were no special classes the writers had to rely on the less certain phenomena of acceleration and retardation as a definite pedagogical expression, in such a system, for subnormality and brilliancy. It was the intention of the writers to consider only such pupils as were sufficiently retarded or accelerated to represent distinct deviations from the average in mental ability. Only those children accelerated or retarded a year or more (by the standard already mentioned) were therefore studied. When the scores were tabulated, the simple expedient was adopted of marking these retards into the score sheets with blue ink, the accelerates with red ink, while the other children were checked in with pencil. A picture of the relation of retardation and acceleration to group-test score was thus obtained, with practically no work beyond what might have been involved in a straight tabulation of the distributions.

It is self-evident that, since the examination was not given below the third grade nor beyond high school, there were no 8, 9, and 10-year-old retards (since these would be in the first and second grades) and no accelerates above fifteen (since these would be through high school). It was finally decided to consider only the retards from 11 through 15 and

[6] It might be added that the product-moments corellation of mental age with group-test score is .63 for this class.

accelerates from 10 through 14, because for these ages only was there sufficient opportunity for acceleration and retardation to show itself. Among the 579 cases in these five ages there were 32 cases retarded over a year, of whom 23 are in the lowest ten per cent for their ages on the total distribution. And there are, in a total of 571 cases, 47 accelerates of whom 18 are in the highest ten per cent.

It is natural that the scale should select more accurately for retards than for accelerates, since a bright child may very well do poorer work in the tests because of chance circumstances of the day or the examination, but a dull child is much less likely, as a result of fortuitous circumstances, to excell himself. The differentiation of neither group is as clear-cut as might perhaps be wished, But from another standpoint the interest of the situation is just here, and the value of such a method of dealing with results is particularly great. To take an actual case —the boy who makes the best score, of the 133 fourteen-year-olds tested, is in the grade normal for his age. His parents wish him to go to work as soon as the law permits, and the boy himself has lost all interest in his schooling. It is surely of no little help in dealing with the situation, to be able to point out that the boy shows ability on the tests above that of children of his own age who are two or three grades ahead of him in school, that he rates above the average for the senior year in high school and yet is idling away his time in the eight grade. In short, there is direct evidence to show that if he would arouse himself to the value of his school work and continue farther in his studies his efforts would undoubtedly be successful and his progress rapid. The highest score, among the eight year children, was made by a boy in the third grade. In urging that he be advanced (as his teacher admits he should be) it is surely an aid in the argument to be able to show that the child makes exactly the same score as another eight-year old boy in Louisville, who is now in the fourth grade and successfully doing two years work in one, and that in his own school there are eight-year-old children one and two grades ahead of him, who make poorer ratings than he. He rated, in fact, above the average child in the fifth grade.

So much for the evaluation of the total scale for the two purposes above mentioned, for separating out feeble-minded cases of institutional grade from the total school population, and for marking off extremes of ability in that population. Some evaluation of the individual tests of the scale remains to be made. As was stated in the previous article already referred to, the tests used in the scale were selected

with some care on the basis of their ability to differentiate dgrees of ability—ability being estimated on the basis of Binet examinations, teachers' rankings, school records and other similar data. But these estimates were made on a relatively small amount of data, and it seemed interesting to make a new evaluation. Such an evaluation was also called for by certain important failings that had already appeared in the total scale. The scale took three-quarters of an hour to an hour to give. This is too long for ordinary school work; provided a sufficient reliability can be obtained, such an examination should not take over thirty minutes, the average school period. An elimination of some tests was thus suggested. Certain tests had already proven relatively unsatisfactory, so far as ease and standardization in giving and ready scoring were concerned. However, if the scale were to be cut down by the elimination of certain tests, the question as to which tests should be eliminated ought to be settled primarily by a measure of the comparative differential value of the separate tests. The method adopted for doing this, and the results, may be indicated very briefly.

III. *The Comparative Differential Value of the Separate Tests of the Scale*

A rough indication of the contribution of each test to the differentiations above mentioned was obtained in the process of working out norms by test. The scores on each test, at each age, were tabulated for the total Bloomington group. This was very readily done, since each test has twenty items and the tests were graded, in the first place, roughly, so that at each age about the same score would be made by the average child on each test. A separate sheet of cross-hatch paper was used for each age. The possible scores (1-20) were numbered across the top; the tests from 1 to 10 down the left hand side. If a nine-year-old child made scores on the ten tests of 9, 10, 9, 8, 6, 9, 4, 3, 8, 5 respectively, his scores were recorded by making a mark at 9 in the space allowed for the first test on the nine-year chart, a mark at ten in the space allowed for the second test, etc. But before tabulating a given age the papers were arranged in order of total score for that age; the records of those scoring in the lowest 10 per cent were then tabulated by test—in black ink. The records of the middle 80 per cent were tabulated in pencil; and finally the scores of the highest 10 per cent were recorded in red ink. This gave on the tables a rough picture of the agreement of

each test with the results by total score in separating off the highest and lowest 10 per cents of the distribution at each age. If test 1 were completely differential of the lowest 10 per cent of mental ability, as defined by the total scale, then the black marks would all appear at the bottom of the distributions for test one, below the lowest pencil mark. Such a test would be 100 per cent differential, or there would be 0 per cent of error. If some of the black marks appeared above some of the pencil marks the test would be less than 100 per cent differential or there would be a per cent of error. If, for instance, on test one at year nine a black mark appears above eight pencil marks, we may say that on this particular case the first test was eight cases in error. And the average of all such errors, on each test, expressed as a per cent of error, will give a ready means of measuring roughly the correlation of each test with the total score at the extremes of the distribution.[7] The average per cent of error of each test, for ages eight through sixteen combined, for the 986 Bloomington school children, runs as follows:

Test	1	2	3	4	5	6	7	8	9	10
Lowest 10.%.....	17	7	11	7	11	8	8	6	11	7
Highest 10%.....	13	8	11	8	17	12	7	10	12	7
Total Error.....	30	15	22	15	28	20	15	16	23	14

The first point to notice is that the tests which are most differential of subnormality are those also most differential of brilliancy. With one exception, the four tests most differential of the one group are most differential of the other, and, again with one exception, the least differential tests are the same. The two errors have, therefore, been combined to make a total error in differentiating the extremes. The two least valuable tests appear to be the first and the fifth. The first is a test of rote memory (the children are read lists of words such as "chalk, man, bag, book, sister" and then asked to write the word that came after "chalk"); the fifth is a test of logical memory (the class is read a newspaper passage and then asked questions about it). The most differential test is the tenth, next come the second, fourth and seventh. The tenth test is "Analogies," the fourth "Opposites." The second is "Logical Selection" (the test contains such lists as "CHAIR —arm, cushion, legs, rockers, seat" and the children are told to "underline the two things, in each list, that the thing named

[7] As may readily be seen, the caluclation of the error can be made extremely simple and rapid; the whole method is much more complicated in the description than in practice.

in big letters at the beginning of the list is *never* without."). The seventh test is "Moral Classification." (The test consists of a list of twenty words, beginning "whispering, setaling, charity, drunkenness," and the children are told to "put a '1' under words meaning something you mustn't do in school, a '2' under words meaning being good to other people, a '3' under words meaning something that is bad for yourself, and a '4' under words that mean getting things from other people in ways you shouldn't.") The fifth most differential test is the eighth, "Dissected Sentences."

It was decided to make a brief scale out of these five most differential tests. It was found on closer study, however, that the seventh test was differential only with the older children; the test permitted sufficient chance successes to blurr the results with the young children. The next test in value, the sixth, was therefore substituted ("Word Completion"—the children are told to fill in the missing leters in such words as "a .it .me .ic"). The other four tests excluded from the brief scale were the "Practical Arithmetic" test (the children were read such problems as "If you had 25 cents, and bought two erasers at 3 cents each and a bottle of ink at 5 cents, how much would you have left?"), and the ninth test ("Practical Information"—the test contains such lists as "bill, coins, check, draft, stamps," the children being asked to "put a line under the safest way to send money, and a cross after the least safe way"), and the two memory tests already referred to. It should be said that the above findings with regard to differential value were not the only factors in bringing about the exclusion of these tests. The arithmetic test was too long, and it made too many demands upon the skill of the examiner, to be easily kept constant from one examiner to another. This fault also held for the two memory tests; besides, they called for writing, and so required more time in the giving and were harder to score. The five tests finally included in the "half scale" were thus not only the most differential, but the best constructed tests as well; both criteria were used in making up the new group.

IV. *The Differential Value of the "Half Scale"*

The half scale was, then, made up of the following tests: "Logical Selection," "Opposites," "Word Completion," "Dissected Sentences," "Analogies." These are, it happens, the even numbered tests of the whole scale. It was now a very easy matter to add up the scores on these five tests, on the records from the full scale, and tabulate the distribution of

half scale scores, as we might expect those results to fall if these tests had been given separately, as a short scale.[8]

The question will be asked, whether the full scale did or did not emphasize, in its sampling of mental functions, some rather special trait, and whether a half scale made of the tests correlating most closely with the scale as a whole may not be merely a scale testing more specifically this one trait? That is, has there not been a harmful "inbreeding"? To obtain some light on this question, the writers have simply evaluated the half scale in terms of its efficiency for the two practical purposes already mentioned, in exactly the same fashion as was done with the full scale.

TABLE I.

HALF SCALE: RESULTS FROM 1022 SCHOOL CHILDREN IN BLOOMINGTON

Age	No.	10%	M.	90%	Grade	No.	10%	M.	90%
8	53	15.3	31.5	44.8	3	116	9.9	27.6	42.8
9	92	15.5	33.2	55.0	4	99	16.4	33.3	48.0
10	96	20.8	46.5	66.8	5	96	33.6	50.1	62.9
11	114	23.5	55.9	74.9	6	108	46.4	58.9	70.9
12	117	41.9	59.5	77.7	7	123	50.4	65.5	76.7
13	111	47.7	65.4	78.0	8	122	59.5	69.6	81.3
14	133	51.6	70.4	85.3	I	123	63.0	76.5	86.9
15	104	55.7	76.2	88.0	II	97	70.0	80.4	88.5
16	89	66.0	79.9	89.6	III	82	71.1	81.0	91.9
17	80	68.5	81.0	90.0	IV	56	77.4	84.5	90.3
	989					1022			

The results can be given in a paragraph. When the scores on the half scale were worked out for all the cases already mentioned, it appeared that for the first purpose mentioned, for selecting out feeble-minded of institutional grade, the two scales gave largely similar results; both scales show three Fort Wayne cases eighteen years or under rating above the 10 percentiles for the Bloomington school children, five rating above the Bloomington five percentiles. But for the second purpose, for aid in determining those children so subnormal or so brilliant mentally as to deserve special educational treatment, the half scale would appear to be, if anything, superior. The whole scale shows 20 of the 23 in the gifted class at Louisville in the upper ten per cent of the Bloomington group;

[8] We do not know, of course, that these tests would behave in the same way by themselves as when given together with the other tests. But norms thus derived may surely be expected to approximate with fair closeness the results which would have been obtained if the five tests had been given alone at least with a closeness sufficient for the purposes of the present paper. The norms thus derived are given in table I.

the half scale places 21 there. The full scale shows 22 of the 25 in the Louisville class for subnormals, the half scale 23, in the Bloomington lower ten per cent. The whole scale places 23 of the 32 Bloomington retards in the lower ten per cent for their age, the half scale puts 26 there; by the whole scale 18 of the 47 accelerates are above the ninety percentiles, 28 between ninety percentile and median, 1 below the median, by the half scale 31 are in the upper ten per cent, 16 between ninety percentile and median, none below the median.[9]

V. *Discussion*

It would seem, then, that for most purposes these five tests would be quite as efficient as the full ten of the earlier scale. They are suggested as a very convenient group for rapid surveys of schools or school systems. The tests take about twenty-five minutes to give, can be printed (or mimeographed) on two sides of a single sheet of typewriter paper, and are readily scored. The norms given above may be thought of as fairly accurate. The half scale may very likely not give so stable a measure, in individual cases, as the full scale; and it gives data much less interesting for research purposes. However, for many practical uses the five tests should prove much more satisfactory.

But the writers have been much more interested in certain general principles involved in the present study, than in any particular results which might be obtained. We very much need more intimate knowledge as to just how important the qualities tested by our tests of "general intelligence" are, in the total complex of qualities that determines social inadequacy and irresponsibility. The institution surveyed contained a much smaller proportion of "Problem cases" than is found in many such institutions, at least so it seemed to the writers. But for all that, the "children" at Fort Wayne were a very heterogeneous group, who required institutional care for a variety of reasons. The writers believe that a systematic study, beginning with a survey by means of group tests, of the total population of such an institution, would yield results of unusual psychological—and sociological—interest. Until such studies have been made we cannot know all that we should about the meaning of our test findings, when we are using our tests to aid in diagnosing "feeble-mindedness."

[9] Expressed in terms of the "average error" above referred to, the whole scale shows 5% of error, the half scale 0.9%, in differentiating the gifted class; errors are 5% and 3% in differentiating the subnormal class. And they are 6% as compared with 3%, and 5% as compared with 2% in differentiating retards and accelerates respectively.

The checking up of our tests by study of groups of exceptional school children who have been independently selected is also worth while. Such methods of evaluation involve errors of their own, which cannot be discussed here; the Fort Wayne cases may have been "institutionalized," and the special school groups given a special training, to an extent which may have affected the results. But at least a rough *indication* has been, in the present instance, obtained of the significance of findings on the scale used, in terms of social inadequacy and of markedly superior and inferior educability. The results have not, perhaps, been altogether satisfactory so far as the merits of the tests are concerned. But at least (as the writers have especially tried to show) so far as the measures *are* reliable they have a much more direct practical implication than would otherwise be the case and it must always be born in mind that special groups may not be, by any means as well selected as they should be; it would certainly be quite as interesting to discover in a special class or an institution for the feeble-minded, a case which did not belong there, as to find among average individuals an abnormal case.

The present study has thus involved a special effort to relate the test findings closely to the practical problems in connection with which they were to be used. The writers cannot escape the feeling that mental measurement has been too much dominated by the concept of "general intelligence"—that the varied educational and sociological problems dealt with in such work are much more complex than is sometimes realized, and that special methods are in most instances needed for each problem.[10] A more intimate contact with the total situation in each special field seems necessary. The writers hope shortly to present data from a number of such special groups (including three institutions for the feeble-minded) with analysis by test of the differential findings.

Summary

1. The "Group Point Scale" distinguishes the children in a state institution for the feeble-minded from a large group of

[10] There was some evidence in the present study to show that the tests most satisfactory in marking off the extremes of educability among school children were not the best for diagnosis of that type of subnormal mentality which requires institutional care. That is, the two tests most differential of the Fort Wayne group were only seventh and ninth in differentiation of the backward school children; the two tests best separating out this group were fifth and poorest in diagnosis of the institutional cases.

school children with only a slight overlapping, in the lower ten per cent of the distribution for the school population.

2. The scale places 42 out of 48 children in a class for brilliant and a class for subnormal children, in the two extreme ten per cents of the total distribution for the school children.

3. It separates nearly as accurately extreme retards and markedly accelerated children from the average group.

4. There is some evidence to show that a scale made of the five tests most differential among the ten used, would be slightly more differential of the groups above mentioned than the entire ten tests.

THE FIELD OF CLINICAL PSYCHOLOGY AS AN APPLIED SCIENCE

A SYMPOSIUM

(The three following papers were prepared for a symposium which was to be presented at the annual meeting of the newly organized American Association of Clinical Psychologists, held in Baltimore, December 28. Owing to the lateness of the hour it was necessary to omit the reading of the papers. Following the suggestion of several members, the papers are published here, without alteration, in order that the material may be made available to the members before the next meeting. J. E. W. W.)

ARNOLD GESELL, Ph. D., M. D., Yale University

The purpose of this symposium is to direct discussion. It is hardly necessary to simply provoke discussion. We are dealing with a very unsettled subject; and it is rather too easy to call forth difference of opinion. Two considerations, however, are no longer in the arena of debate; first, that there is a specialty entitled to the name of clinical psychology; and secondly, that there should now be some form of organization to protect and to promote that specialty as a profession.

When, however, we bluntly raise the question, "What is clinical psychology?" consensus of opinion by no means prevails. Some persons object to the term altogether, and prefer such designations as consulting psychology and applied psychology. Both of these substitute terms, however, are more general and less descriptive. Specifically, clinical psychology is the science and art of individual mental examination and interpretation. It is rather more an art than a science, because its objective is the determination of the mental status of a subject, and the deduction of a practical conclusion as to the possibilities and limits of his improvement. A clinical psychologist is one who is qualified to make a responsible, interpretive mental diagnosis.

Clinical psychology is to be distinguished from psycho-technology. Psycho-technological work has to deal with special and more or less technical questions of methodology and procedure in the fields of industry, commerce, advertising, salesmanship and school administration. Educational psychology, the mental survey, group testing and even trade testing have more affinities with psycho-technology than with clinical psy-

chology. The latter has to do with diagnostic individual examination.

In actual practice the field of clinical psychology has usually not attempted to include individuals suffering from definite mental disease. There are sound reasons for this. There are many problems common to psychiatry and to clinical psychology, but there is also a cleavage which must be respected. There is a considerable field of psycho-pathology which legally, traditionally, administratively, and even theoretically is intimately part of the science and art of medicine. Psychiatry is a recognized branch of medicine; while clinical psychology must, as yet, be considered in the relation of an auxiliary, consultative specialty.

Important work lies ahead in bringing the medical and psychological professions into mutual association; and perhaps the organization which we contemplate may further that work. There is another field of medicine where cooperative relations remain to be established. This is in the important branch of Pediatrics. The whole pre-school period from birth registration to entrance upon school life is destined to become of increasing importance from the standpoint of public health, of public education, and of human engineering. The activities in this field at present are largely medical in their character and in their administration. Pediatrics as a division of clinical medicine or in a broader sense, as a division of public health holds tremendous possibilities. These should in some way be made to include the methods and spirit of clinical psychology. An alliance of some kind must be worked out, or there must be an increase of the type of pediatrician whose equipment includes clinical psychology.

Should not courses in mental development, clinical psychology and mental hygiene become an accepted part of a medical education, and particularly part of the training of the school physician and the pediatrician? And, should not our organization further a movement in this direction? It should be the function of the association to promote clinical psychology as well as clinical psychologists. Nothing will so widely diffuse the benefits of clincial psychology as a proper coordination with the education of medical students, of public health nurses, school nurses, and medical inspectors.

"Clinician" is one of the most respected and significant words in the domain of medicine. It will be fortunate if this word loses none of its dignity and flavor when transferred to the field of applied psychology. A good clinician, in medicine, is an expert who, without necessarily being an authority or

even an investigator in any particular phase of medicine, has so mastered the pertinent principles of his medical science that he can, in the presence of perplexing ambiguities and complications, make a shrewd estimate regarding the conditions and outlook of his patient. He can summon a large amount of previous experience to bear upon his problem. By virtue of his trained perceptions and his native good judgment he can draw a conclusion which has professional and practical value.

Clinical psychology has not yet become as refined, as exact, nor as complex as clinical medicine; but if it is to be worthy of a similar position in the field of applied science, it must protect its standards. Quackery and charlatanism should be exposed and opposed; and methods must be found for guaranteeing and certifying proficiency.

The establishment of standards is the most concrete and essential function which an American Association of Clinical Psychologists can fulfill. This fact at once raises a question in regard to our relation to the American Psychological Association. The latter association, contemporaneously with the inception of the A. A. C. P., appointed a Committee on Qualifications for Psychological Examiners and other Psychological Experts, with a Sub-committee on Certification. Certainly this creates a situation which is in need of adjustment. It is by no means an ominous situation, for practically all of the members on the committee of the A. P. A. are also charter members of the A. A. C. P. There should, however, be a prompt readjustment so that there will be no conflict or duplication of activities. One solution of the situation is to convert the impending A. A. C. P. into a sub-section of the A. P. A. The writer would prefer two discrete organizations with an official affiliation. This would preserve the close relations with the parent organization which are highly desirable, but leave sufficient freedom for prosecuting the ends for which the A. A. C. P. was designed.

Whether evidence of expertness in clinical psychology is to depend upon certification, or upon membership in the A. A. C. P. is after all mainly a matter of detail and administrative policy. We are in no way anxious to create a new academy of immortals, but our aim is to erect high standards for a branch of Applied Psychology and for a possible profession.

HENRY H. GODDARD, Bureau of Juvenile Research, Columbus, Ohio

One of the original Binet tests was "What happens when two people discuss a subject without understanding the words they use?" Most of us know what happens, and yet we are in great danger right now of committing the folly in spite of its known consequences.

Some of us who think we are clinical psychologists think we want an Association of Clinical Psychologists. If it were an informal affair the task would be easy — we would simply pick out the group of people we would like to have associated with us and the thing would be done. If we only wanted a neighbors' club this would do. But we are looking for something more formal, an association that would stand for something. An association to be a member of which is not only an honor but a certificate of standing. Our membership, therefore, must be based upon something more than personal preferences. Whether we organize an association of clinical psychologists or whether we organize a clinical division of the American Psychological Association in either case the members must be clinical psychologists.

What is a clinical psychologist? There is the rub — nobody knows. And nobody knows because the expression is a new one, relatively, and we have not agreed upon what we wish it to connote. The first thing to be settled, therefore, is — what do we wish the term to include? Not what does it include, because as already stated it is a new expression and includes nothing. As everybody knows *clinical* is a medical term which we are taking over into applied psychology, and since we are straining its meaning somewhat in applying it to psychology at all we can stretch it as much as we see fit and make the term mean whatever we like. It will, of course, get its meaning from the group to which it is applied.

Let us first see what it is that we are using. *Clinic* as a noun applies to "the teaching of medicine and surgery practically, at the bed-side or in the presence of patients." *Clinical* is an adjective; "of or pertaining to a sick-bed." A *clinician* "is a physician whose method of treatment, observation, etc., are being based on clinical work. One who holds clinics or whose specialty is the giving of clinical instruction." From these definitions it seems that there are three ideas comprised by the term: first, bed-side; second, sickness; third, teaching. Medical usage has already freed itself from the strict bed-side idea as witness the out-patient clinics. Also they escape more or less from the teaching idea. These same out-patient clinics

very often have nothing of the teaching side connected with them, nevertheless, the physician in his office who sees patients and examines them during office hours does not call that a clinic. The one thing that does remain is the idea of sickness. Physical examination of well people is not called a clinic.

Returning now to the psychological use of the term we know that about all that is left if we consider all the ways in which the term has been used, is the idea of personal examination. Indeed if we should ever come to call the person whose only training and experience is in giving group tests, a clinical psychologist we would have thrown away the last vestige of the original significance of the term clinical. The teaching idea does not hold. If the time ever comes when one teaches psychology by means of clinics they will undoubtedly be called psychological clinics, but that is insignificant at present. The only thing left is the sickness side. Shall we confine its usage to something that is at least analogous to sickness? or, shall we break away altogether and allow the term to be used for any mental examination? It seems to the writer that the latter procedure would be a mistake and that much of the present confusion as to what shall be the qualifications of membership in the proposed association of Clinical Psychologists arises from the extension of the term clinical psychology until it is practically synonymous with applied psychology.

Would it not simplify the whole situation if we were to agree that as related to the practice of psychology, clinical should mean personal examination of some one who is mentally abnormal, or subnormal, leaving those psychologists, who apply the science to determining what the special traits of a normal individual may fit him for, to form a separate group, possibly to be called vocational psychologists? And again those who make a specialty of applying the laws of mind to industrial efficiency to constitute another group, the industrial psychologists? Surely these are three quite different fields. The clinical psychologist as thus defined need know nothing of vocational guidance or industrial psychology, any more than a physician needs to be a clergyman or a lawyer. Or, if the illustration is better, any more than the physician need to be an expert surgeon or an oculist. If this view could be accepted it would indeed go a long way toward settling the other problems that we are facing whether the clinical psychologists will form an independent association or ask to be set off as a division of the A. P. A.

The number of clinical psychologists, under this definition, is as yet too small to warrant the establishment of a separate organization. We are too few to obtain the purposes which we

have declared as those of such an organization. Nevertheless it is highly desirable that the clinical psychologists be recognized, that standards of training and experience be established before the group becomes so large as to be unwieldly and before it includes persons that do not have qualifications that are recognized as essential. In accordance with this plan the A. P. A. might at once divide itself into at least five groups or more if needed: the pure psychologists, educational psychologists, clinical psychologists, vocational psychologists and industrial psychologists. Each of these would in time perhaps, if not at once, subdivide again; for example, clinical psychology might divide itself into perhaps three groups: those devoted to the study and diagnosis (1) of mental arrest, (2) of mental disease, and (3) of crime and delinquency. Of course these somewhat overlap and run into each other, but that is true of all specialization and only means that those who devote themselves to crime and delinquency must be clinical psychologists in a broad sense to the extent that they recognize mental disease and mental arrest. This seems like a logical development of the A. P. A.

If, on the other hand, we make the term *clinical* synonymous with *applied* and form an idependent association of clinical or applied psychologists there will inevitably be sooner or later a breaking up of the present membership of the A. P. A. into two groups. While we would attempt to keep up our intrest in both associations it would inevitably result in a division.

The result of the work of the psychologists in the war has demonstrated clearly that psychology has a practical side and the people are conscious of that fact and are going to more and more insist that psychologists help them solve their problems. This means that the field of applied psychology is as wide as human activity and we will more and more devote our knowledge of the science of mind to the practical solution of human problems. Psychology is today able to do more for human efficiency than perhaps any other science. It is possible to take a child and after a careful clinical examination make a diagnosis which in thousands of cases saves the wasted energy of years in the old trial and failure method of education and industrial activities; and yet we have made only the merest beginning. The work is destined to spread very rapidly and cover an enormous field and if done rightly and wisely will perform an enormous service to humanity.

In the nature of things, however, there will be enormous abuses. Just as the public was for years preyed upon by medical quacks and patent medicine venders, so there will be innu-

merable charlatans practicing under the name of psychologists. This is inevitable. There is probably no way to prevent it but much of the evil can be headed off by vigorous action on the part of an organization of bona fide psychologists. It would seem that the American Psychological Association with its more than a quarter of a century of work, its carefully selected membership has a tremendous advantage, in carrying out this work, over any separate organization that would be formed.

Moreover, we must not overlook the possibility of organizations in the bastard group who will go into this work purely for financial reasons, moreover there is nothing to prevent them from taking high sounding names that will attract the public and make the work of the true psychologist so much more difficult. In view of all these considerations I question whether we can at this time afford to run the risk of dividing our forces by the organization of a separate association of clinical psychologists, whether we use the term in the limited sense for which I have spoken or whether we use it in the broader sense as practically synonymous with applied psychologist. As members of a division of the A. P. A. we would all be bound together with strong *esprit de corps;* we would hold our meetings together and would have an ever increasing influence and power with the American public.

J. E. WALLACE WALLIN, Psycho-Educational Clinic and Special Schools, St. Louis.

In the field of medicine the connotation of the word "clinical" is perfectly well fixed. The clinical method refers to the direct examination of the individual patient by any or all of the approved methods of modern medical science. The aim of the examination is: first, to classify or diagnose the disorder or disease upon the basis of the etiology, which frequently can only be done by a systematic, comprehensive examination of the individual; second, to prescribe a course of curative treatment; and, third, to make a prediction of the probable outcome of the treatment and the future course of the malady. Of course, the physician may examine normal, healthy individuals as well as those who are ill, but in practice the time of the vast majority of medical men is given to the examination and treatment of the diseased and disabled. At the present time we note a growing tendency to place the chief emphasis upon the preventive aspect of medicine, and the interest in preventive medicine has already led to the development of a new profession, and the establishment of a new degree, the Doctor of Public Health. This type of doctor is not primarily, if at all,

engaged in the work of clinical examination, but is concerned with maintaining the public health by the control and regulation of the sanitary agencies affecting the well being of the entire community.

To prepare competent clinical practitioners the medical schools have established courses of instruction only part of which are clinical in nature, but all of which serve to prepare the student for the strictly clinical branches in the upper years of the course. The degree of M. D. is not conferred upon anyone who has not completed all of the preparatory and clinical courses. In addition eight medical schools now require at least one year of clinical practice in a hospital internship before conferring the degree. No one is admitted to membership in the largest and strongest organization in the medical profession, the American Medical Association, who is not a "legally registered physician." In only a few states is it possible to register as a physician without the M. D. degree. To still further safeguard the interests of the public, no one is allowed to practice medicine who has not received a state license.

Now, I take it that the situation in clinical psychology, with reference to the significance, purpose, demands and responsibilities of psycho-clinical work, does not differ materially from the clinical situation in medicine, at least so far as concerns work with the mentally disordered or deviated. Let us now briefly emphasize a number of propositions which are germane to a consideration of the field of the clinical psychologist and the plan of organization of this Association.

First, no matter how valuable the results from group psychological tests may be for the relative rating of the members of a group or for securing preliminary data, it is questionable whether we are justified in considering group testing as clinical testing, and classifying group testers as clinical examiners. The clinical examiner may employ group tests, but he is not a clinical examiner unless he analyzes (or diagnoses) each case on the basis of an individual examination. A clinical examination in psychology, as in medicine, always means an individual examination of a subject.

Second, the field of clinical psychology in the past has, in practice, been largely restricted to the examination of mentally deficient, backward and delinquent subjects, and possibly the psychotic, although the workers among the insane have usually not classed themselves as clinical psychologists. Theoretically, however, the field of clinical psychology is coextensive with the field of individual psychological examination. The clinical psychologist may devote himself to the study of the feeble-

minded, backward, normal, supernormal, the delinquent, the dependent, the pauper, the hobo, the psychotic, the specifically mentally handicapped, the applicant for vocational guidance, the person who wishes to know whether he possesses aptitude for a special craft, art or profession, the army and navy recruit, with a view to insure placement in the proper arm of the fighting forces, etc. But can we classify any investigator of these classes of individuals and of these problems, no matter how meritorious his work, as a clinical psychologist unless he utilizes the clinical method of individual examination? Testing army recruits in groups of hundreds by means of intelligence test-blanks doubtless under favorable conditions gives a fairly reliable and valuable relative intelligence rating, but such an examination is no more clinical than a class examination in geography or reading, and such a group tester is no more to be considered a clinical examiner than a teacher who conducts class tests in the school branches. When the clinical method is used, however, there is theoretically no limit to the psychological problems which may come within the scope of clinical psychology.

Third, the purpose of a clinical examination in psychology, just as in medicine, is: (a) the making of a correct diagnosis of the case, which involves the elaboration of a clinical picture of the mental conditions found at the time of the examination, and the detection of the etiological factors responsible for the ascertained mental deviations; (b) the rendering of the advice, or the prescription of the educational and social treatment which is indicated by the diagnosis, together with the reinforcement of recommendations for physical treatment made by competent medical examiners; and (c) the making of a reliable prognosis, so that the treatment or training may be effectively directed toward the best development of the individual. Whoever fails to realize all of these aims only partially performs the functions of a clinical psychologist. Whoever has been trained to do less than this, has been inadequately trained for clinical work in psychology. Certainly so far as concerns the mentally pathological or abnormal or anormal classes, the clinical psychologist, just like the physician, must deal with etiological factors and not merely with symptoms. Just as the physiologist or the anatomist, no matter how much he may know about the human body, is not a physician because he does not employ the complete clinical procedure, in the use of which he has no special fitness, so the general, experimental, educational, or physiological psychologist is not a clinical psychologist, no matter how much he may know about the mind, unless he has

been specially trained to use the clinical procedure. Therefore—

Fourth, to become a competent clinical psychologist, just as to become a competent examining physician, requires a course of special training. A course of training in the ordinary branches of psychology which is entirely adequate to produce a good teacher or investigator in those branches is not sufficient to produce an efficient clinical examiner in psychology. But the training needed in addition to the basic psychological, biological and scientific training will differ according to the field in which the clinical psychologist is engaged, and will have to be gradually determined by the consensus of opinion of the workers engaged in each specialized field. The strictly professional training — *i. e.*, the training beyond the basic psychological and scientific training — which will qualify a psychologist for skilled work among the psychotic, will not qualify him to diagnose and direct the educational treatment of feeble-minded, backward, speech defective, alexia — or dyslexia — or merely pedagogically handicapped school children. Nor will the training which qualifies one for efficient work among these types of handicapped school children produce an expert on the selection of successful salesman or on the proper placement of army recruits, or on the direction of courses of instruction for supernormal children. The different fields of clinical psychology have their own special and unique problems, the proper handling of which requires special study and skill in the particular field concerned. Doubtless the technical differentiation of psychological specialists will so develop that in future we shall have some psychologists who will specialize on feeble-minded children and adults, others will specialize on backward and specifically handicapped school children, others on delinquents, others on bright and talented children, others on the selection of employes, etc.

This is not the place in which to attempt to outline the kind of differential training which clinical psychologists working in different fields should have. But we may indicate briefly, by way of illustration, the training which we believe workers must have who would render skilled service in the psychological examination of mentally and educationally handicapped school children, and in the supervision of educational work in their behalf.[1]

[1] For a further discussion of this question, see our article on The Field of the Clinical Psychologist and the Kind of Training Needed by the Psychological Examiner of Mentally and Educationally Deviating School Children. *School and Society*, 1919, p. —. See also *The Mental Health of the School Child*, 1914, pp. 114f, 134, 136, 210, 216f.

1. The basic training in the various branches of psychology.

2. Special training in clinical psychology, including training in the administration of the current tests used in individual examination, the practical clinical examination and the working up of case reports of at least 200 cases, in course, and preferably a post-doctorial year in a position corresponding to a medical internship, in which studies would be made of various types of deviating children, especially the feeble-minded, backward, epileptic and psychotic.

3. Various pedagogical courses, particularly courses on standardized educational tests; on primary methods of teaching, especially the teaching of reading, spelling and number, and the kindergarten and sensori-motor exercises; on the psychology and pedagogy of the feeble-minded, backward, and speech defective, and children with specific sensory defects, particularly visual and auditory, and with specific defects in reading, writing, spelling and number; and practical courses in the various types of handwork which should be offered mental retardates.

4. Social pathology, including a study of the social, vocational and criminal aspects of mental deficiency and defect.

5. A minimum amount of medical training, in order that the clinical psychologist may secure proper perspective in the work and be able properly to interpret the medical reports, including courses in physical diagnosis, pediatrics, orthopedics, ear, nose and throat disorders, mental deficiency from the physical point of view, neurology and psychiatry, and practical case writing, including the study of the subject's developmental and family history.

It must be quite evident that while the above course of training would qualify a psychologist for skilled service with educational deviates — and we doubt that a less thorough course would do this [2] — it would not qualify him for expert service in many other fields of clinical psychology. We doubt that anyone would propose the above course of instruction in its entirety for a student who was preparing to make psychological studies among the insane, or who was qualifying as an expert on trade tests or on the selection of salesmen.

Fifth, owing to the important practical bearings of clinical work in psychology, only those should be accorded official or

[2] How far the psychological examiners in the public schools of the country fall short of a lower standard of preparation than the one suggested above, was indicated by the study of our questionnaire returns from 302 cities in 1913. In our judgment only about 25% of 115 psychological examiners at that time were qualified for the work. Cf. *The Mental Health of the School Child*, 1914, 392f.

professional recognition as clinical psychologists who have been adequately prepared. The organization of an association or associations, with rigid eligibility requirements, will do much to bring about the proper training of examiners and secure for them the recognition to which their professional qualifications entitle them.

If the above conclusions are correct, the following corollaries would seem to follow. We do not submit them in a dogmatic spirit, but rather as suggestions which we think should be adequately considered before a constitution is adopted by this Association.

First, no one should be eligible for membership in the American Association of Clinical Psychologists unless he is at the time of election holding a position in which he is regularly devoting at least a portion of his time to actual psycho-clinical examinations, or unless he has in the past devoted a considerable portion of his time to genuine clinical work. Possibly it were wise to make an exception in the case of teachers who are training students in mental testing and of research workers who are perfecting and standardizing psychological tests which will be of value in individual examinations. It is evident, however, that teachers will not be especially qualified to train the students in the art of clinical diagnosis and treatment, which involves more than training the students to administer psychological tests, unless they have had actual, first-hand clinical experience and are devoting some of their time to clinical examination. And yet we must recognize, from what has already been said, that the necessity of having clinical experience and doing clinical work, will vary with the type of problem.

Second, no one should be eligible for membership who has not taken the Ph. D. degree in the various branches of psychology in an institution of recognized standing; and who has not published or prepared for publication a valuable contribution in the field of mental tests or clinical psychology.

To this condition it is probably advisable to make one exception: a holder of an M. D. degree may be admitted if he has had special training in psychology and considerable experience in making individual psychological examinations and diagnoses, and if he has made valuable contributions in the field of mental tests and clinical psychology.

The suggestion has been made that the rank of associate membership be established for psychological technicians or testers who cannot qualify under the above conditions. We are inclined to doubt the advisability of this suggestion. One of the chief aims of the Association should be to develop the

highest standards of work in the field of clinical psychology, in order that properly trained clinical psychologists may secure the emoluments and recognition to which their special preparation and fitness entitle them. We shall probably secure a large number of applications for membership from persons who wish to profit by the prestige which membership will confer who should not be given this recognition if their basic and professional training and experience are inadequate. In this connection, however, we may raise the question whether it would not be advisable to call the members of the Association Fellows rather than Members.

Third, in addition to the above two general conditions, a candidate to be eligible for membership must have secured a special or technical preparation in addition to the basic work in experimental; educational and other branches of psychology. This strictly technical preparation, however, would have to differ, as we have already stated, according to the particular field in which the psychological examiner is engaged. If this is so, it is impossible to outline the detailed professional training which a candidate for membership must have, unless the Association is divided into sections. Every section would then have to determine upon the necessary technical training needed for membership in its own section. Under such an arrangement every candidate would have to meet two requirements: first, the general conditions for membership imposed by the association at large — say, the first two conditions given above — and second, the special conditions required by the individual section. If this plan were approved, we should probably have the following sections: Mental Deviation, including mental retardation and deficiency and possibly specific mental deviations; Mental Disorder, dealing with various grades and types of psychotic disorders; Crime and Delinquency; and Trade and Vocational Testing. Other sections would eventually emerge as the problems become more clearly defined, because the progress of the science will doubtless be toward greater and greater differentiation of special fields. The American Medical Association now has 16 sections in its Scientific Assembly, and in addition five councils which make recommendations in the interest of physicians.

One of the practical arguments for establishing the sections is that this will make it possible to more fully realize the aims of the association to establish high standards of professional qualifications for clinical examiners in psychology. According to the plan of organization as at present proposed, it is impossible to define satisfactorily the strictly technical qualifica-

tions which a clinical psychologist should possess in order that he may be rated as proficient or in order that he may be eligible for membership, because the requirements which are applicable to one type of examiner are not applicable to another type. The Association already has in its membership persons who should be distributed in several sections, while applications in future will no doubt come from persons representing very diverse interests. Recently a western state adopted a statute according to which no one can be appointed as a clinical psychologist to one of its state institutions for the feeble-minded who is not a member of the American Psychological Association. But the majority of members of that Association have not made any technical study of feeble-mindedness at all, so that the fact that one is a member of this Association does not constitute him an expert on feeble-minded children. But the same criticism applies also to this Association. Membership in this Association, according to the present plan, does not indicate whether the person is qualified to do work with feeble-minded or backward school children, or with the insane, or the delinquent or the normal, or any other class. The establishment of a number of sections, with membership limited to the experts in the particular field of each section, in accordance with the present practice of the American Medical Association, would have the merit of indicating the particular specialty of each member. Would it be worth while to divide the Association up into sections to gain this advantage? I believe that it would. But there may be weighty practical objections to sectionalizing the membership. The aim of the writer is merely to offer the suggestion, in order that the whole matter may be fully considered before final action is taken.

If the suggestion were adopted, however, the Association could adopt the general conditions for membership at this meeting, while the specific technical qualifications would have to be worked out by committees of the various sections which would be authorized. These committees would report their recommendations to the individual sections for adoption at the next annual meeting. No new members ought to be elected until the sections had determined upon their eligibility requirements.

A further suggestion has been made that the title of the Association be changed to the American Association of Consulting Psychologists, or the American Association of Applied Psychologists, and that sections be established to include all the important applied divisions of psychology. The scope of such an Association would be broader than that contemplated by the present Association, as non-clinical workers would

evidently be eligible for membership. If eligibility to membership in the Association of Clinical Psychologists will be limited to persons who are engaged in genuine clinical work, at least on a part-time schedule, or have been thus engaged, this expansion of the Association is probably not advisable. But if eligibility will not be so limited, it would seem to the writer that either of the other two titles will more correctly express the scope and purpose of the Association.

BOOK REVIEWS

PAUL POPENOE and ROSWELL H. JOHNSON. *Applied Eugenics.* The MacMillan Company, New York, 1918. p. 443.

This book, while bound to provoke vehement discussion, is as surely bound to afford no small amount of satisfaction to those who have felt acutely the want of a thorough-going treatment of the principles and program of the eugenic movement, since gross misunderstanding and misuse of the term Eugenics still persists.

For the first time we are in possession of a balanced study of adequate length, which escapes being a primer or a profoundly technical volume. It is greatly worth the waiting for.

The lay reader will be impressed with the incisive refutation of popular confused assumptions involved in the old debate upon the relative importance of heredity and environment. The lines of evidence, statistical and experimental, indicating that differences in a trait are more affected by differences in nature than by differences in nurture, are drawn from the Galton, Thorndike, and Pearson investigations and from the researches in families of distinguished persons made by Woods and Cattell. The latter's study, however, only indirectly proves the author's point in that it serves to call attention to the complex interrelation of environmental factors operating upon traits inherent in original nature.

A discriminating analysis of the problem of germ plasm modifiability is provided. Current misconceptions concerning the transmission of characters are summarized and the actual significance of data traditionally alleged to show it is thoroughly pointed out.

Contemporary psychologists and biologists need no new treatise on the distinction between the conditions affecting germ plasm and soma plasm or upon the capacity of natural selection to account for progressive and retrogressive changes in the race. What we did need was an authoritative and at the same time readable presentation of the facts for non-scientific persons, many of whom, though academically trained, are still advocating as *eugenic* various agencies on the false supposition of their fundamental power for racial betterment. This need is met in a book likely to surprise, inform, and convince such individuals if they can be induced to read it.

Furthermore the doctrine regarding "racial poisons" is still tenaciously held by many sociologists, social workers, psychiatrists, and clinical psychologists, unaware of the recent results obtained through close laboratory scrutiny of the subject. These students fall into two classes: those unconscious of the fact that the influence of alcohol, for example, upon the mechanism of heredity is a matter even open to discussion (as two persons who are regarded as experts in the field of social science recently confessed their ignorance to the reviewer); and those who, aware that it is questioned, resent what they deem audacious and preposterous attempts to minimize the baneful effects of alcohol upon the mind and body.

The authors give a critical resumé of the evidence on both sides of the controversy. They further augment the weight of the argument against transmission of acquired characteristics by an admirable crit-

ique of the time-honored belief in prenatal culture. In emphasizing the conclusion that eugenics is obliged to concern itself solely with the material in the germ plasm which cannot be permanently modified by external influences, and that hence the only means to race betterment is through changing the birth and death rates to alter the relative proportions of good and bad germ plasm in the race, the authors are by no means scornful of Euthenics. It is to be expected that they will be wrathfully accused by superficial readers of denying the good accruing to the individual and society through all the reform movements of the day, in spite of numerous explicit statements such as this: "Eugenics . . . is inadequate unless accompanied by constant improvement in the surroundings; and its advocates must demand euthenics as an accompaniment of selection, in order that the opportunity for getting a fair selection may be as free as possible."

The chapters dealing with mental differences and inheritance of mental capacities are well handled from the standpoint of a disciple of the Thorndike school upon whose work the authors draw heavily.

Although their resumé of the literature dealing with the comparison of the white with the negro races is excellent, psychologists will be quick to note the omission of reference to Woodworth's and Bruner's work on *Racial Differences in Mental Traits* and Thorndike's penetrating valuation of the facts. The desirability and possibility of enforcing restriction on marriage and parenthood among the dysgenic classes is argued on biological, humanitarian, and economic grounds. Legislation of genuine eugenic value would be segregation, differently drafted sterilization and divorce laws, and repeal of laws prohibiting marriage between cousins.

In the chapter on Improvement of Sexual Selection the authors state what they call the psychological point of view in support of the claim that eugenic ideals can become a part of the habits and "conscience" of the race and the selection of life partners unconsciously influenced. The authority cited for this belief appears to be Galton alone. Watson's advocacy of experimental study in the control of emotions and the application of the theory of the "conditioned reflex" as a means of reaching such an end has been overlooked.

One of the features of the book which bears the distinction of original contribution is the chapter pointing out the eugenic aspects of certain specific laws, movements, and reforms, e. g., Democracy, Socialism, Feminism, child labor, compulsory education, vocational guidance, minimum wage, trades unionism, housing, pensions, taxation, prohibition, *et cetera*. The authors are not guilty of sweepingly branding these features of society eugenic or dysgenic, but cogently set forth their direct and indirect influences in respect to ultimate racial improvement through increase and decrease in relative proportion of superior and inferior strains in the population. The conclusions reached through adroit reasoning are perhaps more often unexpected than not.

Any consideration of dysgenic effects of social amelioration must take into account that we do not know what the trend of variation in human beings is nor how natural selection is operating. On the other hand, conditions entailing a desirable high death rate of inferiors may at the same time prove unwholesome to the superiors as well and even actually create inferior variations unaware. Until we have experiments showing what environmental forces are the causes of variability we cannot hope unerringly to determine and eliminate the true sources of deterioration.

There are certain infelicities of diction and noticeable typographical errors, but it is none the less an admirable specimen of book-making. The interest of the discussion is enhanced by copious and effective illustration.

Smith College. MIRIAM C. GOULD.

HELEN MAROT. *Creative Impulse in Industry, A Proposition for Educators.* E. P. Dutton & Company, New York, 1918, p. 146.

This book, which carries the significant sub-title, "A Proposition for Educators," is dedicated to "Caroline Pratt whose appreciation of educational factors in the play world of children, intensified for the author the significance of the growth processes in industrial and adult life." The author analyzes the industrial system in this country and by contrasting it with the German system points out its inherent weaknesses as well as the means of remedying them. "America wants industrial efficiency, it must have efficient workers if it holds its place among nations, and American people will prove their efficiency or their inefficiency as they are capable of using the heritage which industrial evolution has given the world. But what shall we use this efficiency for? For the sake of the heritage? For the sake of business? For the sake of Empire?"

Among the chief weaknesses of our present industrial system, according to the author, are the facts that we value possession of wealth more than its creation, that both capital and labor are interested in the exploitation of wealth rather than in its production. Consequently the manufacturers value the workmen merely as so many attachments to their machines which they would like to standardize, perhaps, by an industrial education, just as the other manufacturing processes have been placed on the highest basis of efficiency as measured by mass production. The monotony in labor brought about by thus dividing "the workers into eyes, arms, fingers, legs, the plucking out of some one of his faculties and discarding the rest of the man as valueless," his ignorance of the complete process of production, of the finished product and its final marketing, have produced in the laborer not only a mental state of indifference to his work but also a disinclination to produce, which in the long run may offset such artificial stimulants and inducements to higher efficiency as better wages, comfortable and safer working conditions, means of self-improvement, and selection for positions and advancement to better positions on the basis of greater fitness. It would be useless to revert to and repeat the experiences of craftmanship and the Guilds of an earlier age, since our present state of society has evolved from it. Other suggestions for correcting the evils of our present system are not promising, since they fail to appeal to the worker's impulse of creativeness. The author then goes into some detail, in the second chapter, to show how only partially successful in this respect the efforts of scientific management have been except in so far as they incidentally and accidentally involved this impulse. Another chapter is devoted to the attempt at showing that the German paternal system of industrialism for the glory of the empire would be absolutely impossible for the American people, even if greatly modified and in spite of our recent favorable experiences in governmental control for the purposes of carrying on the war.

The author's own positive suggestions for remedying the most seri-

ous evil of our present industrial system consist in demonstrating the possibilities for developing a socialized productiveness in the workers, by utilizing or appealing to their "orginating, choosing, adapting power." She says, "the creative effort is not *necessarily* an individual matter. It may be possible for a group of people to associate cordially and freely together with a single creative purpose and endeavor. It may be possible for each worker to experience the joy of creative work as he takes part with others in the planning of the work along with the labor of fabrication. It is a creative experience or dull labor as his association with others in the solution of the problem is freely pursued and genuine, or as it is forced and perfunctory" (p. 137). In this connection the author describes in some detail the educational experiment carried on by Caroline Pratt in New York City under the auspices of The Bureau of Educational Experiments, 16 West 18th Street. Miss Caroline Pratt, with the help of several adults, is conducting a Toy Shop with about forty pupils between 14 and 17 years of age. The pupils have here an opportunity to become familiar with all the technical, financial, administrative, economic, aesthetic and literary processes involved in the production of marketable toys for American children. The practical value of such an educational experiment will depend upon the answers that it can make to such questions as "(1) Was the creative impulse aroused? (2) Were standards of workmanship discovered and sustained? (3) Was a broad as well as a working knowledge of subject matter acquired? (4) Did the children approach established methods in a spirit of hospitality and of inquiry as to their validity? (5) Did the problems create sufficient interest to arouse the desire and will to reject faulty methods, and introduce others of greater service? (6) Was the enterprise a productive one from the point of view of the market and an educational one from the point of view of growth?" (p. 133).

The author recommends that educators cooperate with engineers and architects and the managers of industry in proposing a new program or system of production. She believes that "engineers, not under the influence of business, are qualified to open up the creative aspects of production to the workers and to convince them through their own experience that there are adventurous possibilities in industry outside the meager offerings of pay-day" (p. 109). She names Mr. Robert Wolf as one of the engineers who is ready for the adventure. In working out such a program "it will be the mission of educators to show that by opening up the field for creative purpose, fervor for industrial enterprise and good workmanship may be realized, that only as the contents of industry in its administration as well as in the technique of its processes is opened up for experiment and first-hand experience, will a universal impulse for work be awakened" (p. xx).

Whether it may be granted or not that the book correctly diagnoses our present industrial defects, it fails to combat the employers' claim "that from 95 to 99 per cent of the working force is without productive impulse" (p. 134) by the only convincing argument to the contrary, the experimental demonstration of the universality of a creative impulse in human nature. In this respect the author's mere assumption of its existence is psychologically as unfounded as Veblen's assumption of an instinct of workmanship. To recommend sweeping educational and industrial changes on the basis of a new sociological hypothesis, because our present industrial system is faulty, amounts to suggesting the haphazard methods of trial and error instead of the

much slower but surer way of first establishing the psychological facts upon which the new sociological hypothesis must ultimately rest.

L. R. G.

ALICE MINNIE HERTS HENIGER. *The Kingdom of The Child.* (With an Introduction by G. Stanley Hall.) E. P. Dutton & Company, New York, 1918, p. 173.

In her latest volume, Mrs. Heniger presents the psychological conclusions which her fifteen years' work with the Children's Educational Theatre have enabled her to draw, and shows how they may be used as fundamental principles of modern pedagogy. The child's kingdom is the world of imagination, of make-believe, is her main thesis, and we must use this innate creative instinct in order to breathe the glow of life into our deadly school curriculum, and also to mold the character and shape the personality of the children intrusted to our guidance. The vital mistake we have made as parents and teachers is our failure to recogize this basic tendency of childhood, as a result of which we have forced our children to repress their impulses toward self-expression, and made them become servile imitators of us, as adults too blind to see the longings and aspirations of their prisoned souls.

Mrs. Heniger gives numerous examples which illustrate her point more clearly than any abstract discussion could ever hope to do. She cites cases of retarded children, with whom the ordinary school methods have failed, and tells how, under the benign influence of the Educational Theatre, their lives have grown fuller and richer in their strivings to interpret the characters assigned to them, so that the long denied promotion becomes an indirect result of this release of their inner life. She describes the months of arduous study put in by a little slum girl who wished to be able to play the part of a fashionable boarding school miss in "The Little Princess," and pertinently inquires whether any other motive could have induced her to work so eagerly to acquire the traits of personality and social grace of a type far outside her personal experience. For these children, starved by their environment, seek to enter another world vicariously through the gates of dramatic representation; they live the parts they play,—indeed, they must become the person in the story before they can hope to portray the character with any degree of success. Herein we have one solution of the problem of Americanizing our immigrant children; we have only to make them enter into the spirit of our great men,—as one little Russian lad did in his rendition of Lincoln's Gettysburg address,—to produce more loyal citizens than could be obtained by any amount of drill upon the Declaration of Independence as a mere rote memory lesson. It is only as they can enter into the lives of the men who created that glorious document, it is only as they, too, can feel the thrill of freedom and justice, that these foreign-born youngsters are made an integral part of our national life.

In like manner, the subjects we wish to teach our children must become humanized, and must present to the child a many-sided contact with the great racial history of mankind. Literature lends itself particularly well to this interpretation, for under skillful guidance, the child enters into the soul of the heroes and heroines of whom he reads, and unconsciously strives to emulate their noble thoughts and lofty idealism. It is in the adolescent age of the high school and college that we see the student most affected by these inspiring ex-

amples; but even the tedious task of learning to read may be made easy and pleasant if the teacher be wise enough to make the cat a real cat, and the dog a barking, tail-wagging animal of daily acquaintance.

The Sunday School comes in for its share of censorship, and numerous anecdotes are quoted to illustrate the hopeless inefficiency of well-meaning but unintelligent teachers, with their precise but uninspiring presentation of facts. It is worth while to repeat one of these examples, for it drives the point home as nothing else could do. "The story was recounted to me," writes Mrs. Heniger, "of one such lady who thought it would be appealing and helpful to tell her class the story of Adam and Eve in her own language. She had read somewhere about childhood's ready grasp of the objective lesson, so she decided to cut the story down to what she considered the level of the child's understanding. She was greatly pleased with her apparent success inasmuch as the class . . . listened attentively to her recital. But here is the way the brightest lad in the class recounted the story to a number of young friends. I repeat his language verbatim: 'There was two of 'em, and they ate off the wrong tree, and the snake told on 'em, and he said *she* made me do it, and so God said you both go to Hell!'" After driving home her argument by an anthology of these childish observations, the author pleads for an interpretation of the bible which shall bring out all its mystic beauty and allegorical significance, and pleads for the utilization of the dramatic element in the Sunday School, emphasizing the value of biblical plays for the childish actors.

While these details serve to show the style of the book, and the methods which it advocates, they can give no clear idea of the author's keen insight and breadth of vision. The scope of the work is partially indicated by the elaborate chapter headings: What Is the Kingdom of the Child and How Does the Adult Transgress in the Kingdom; How Can the Parent and the Teacher Enter with the Little Child into His Kingdom; Shall Parent and Teacher Be Subjects in the Kingdom of the Older Child or Shall They Be Rulers Therein; In the Kingdom the Child Can Learn Best How to Know Himself by Being for a Time, Someone Else, but He Must Create Not Imitate, His New Character; In Every Child within the Kingdom Slumbers the Whole Experience of the Race and We, as Educators, Must Endow the New Spiritual Lives with Form and with Substance; How the Foreign-Born Child Is Brought into the Kingdom and Remains Forever After a Loyal Subject Therein; Learning to Read Is Never Dull and Tiresome when the Schoolroom Is a Part of the Kingdom; All the Great Heroes and Heroines of Literature Reside in the Kingdom where Children and Young People Crave to Be Their Daily Companions, to Live with Them in Their Homes and to Interpret Their Thoughts; In the Kingdom, Literature is Regarded as an Expression of Our Common Humanity and Speech Is the Direct Medium from Soul to Soul; The People Who, in Books, Seem to Be dead and Buried, Are All Alive when We Meet Them in the Kingdom; In the Bible Are Page after Page of Attractive Stories that Contain Guiding Lessons of Life; In the Kingdom, Bible Stories Are Shown as The Wonderful Symbolic Allegories They Are, and Which Children's Pure Aspiring, Spiritual Imagination Really Imagines Them to Be; The Kingdom's Domain Stretches Beyond the Walls of the University, so that even a Staid College Professor May be a Dweller in the Kingdom if He Has a Mind to; The Real Kingdom Will Have a Children's Theatre

Where Good Plays Suitable for Children to See May Be Enjoyed; In the Kingdom a Little Child Shall Lead Us.

In his introduction to the present volume by Mrs. Heniger, G. Stanley Hall characterizes it as "a real and original contribution to both educational methods and our knowledge of childhood." He might well have added that it is a book which should be in the hands and minds of every mother and teacher, for it contains information which is vital for the proper development of the child's social and moral nature.

PHYLLIS BLANCHARD.

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. III JUNE, 1919 No. 2

AIR SERVICE TESTS OF APTITUDE FOR FLYING[1]

By V. A. C. HENMON, University of Wisconsin

From the psychologist's point of view an outstanding feature of the war has been the demand for highly specialized abilities and aptitudes in addition to general military virtues. No where has this been more true than in the Air Service and no where have the problems of selecting individuals and assigning them to the work for which they are best fitted been more urgent. As late as April 1917, the Aviation Section of the Signal Corps had but 52 trained flyers. The Air Service with upwards of 16,000 flyers is thus a new product.

When war was declared the need for selecting thousands of men for training as pilots, observers and balloonists was met by the creation of Examining Boards and Physical Examining Units. Rigorous physical and medical standards were adopted. While the mental, moral and professional requirements were equally exacting the standards were necessarily vague and general. We were instructed to select men of good education and high character, men who were in every way qualified and fitted to become officers of the U. S. Army — a rather intangible set of specifications. We were constantly enjoined to remember that the flying officer was not to be an "aerial chauffeur," but a "twentieth century cavalry officer mounted on Pegasus."

Examiners soon found that while it was relatively easy to pick men who could negotiate the non-Pegasarian flights demanded in the Ground Schools, what they would do when

[1] Paper prepared for joint meeting of Section H of the A. A. A. S. and American Psychological Association at Baltimore. Delay in securing approval of War Department prevented the reading of the paper.

they mounted Pegasus was a sheer guess. The result was that in spite of the Examining Boards and Physical Examining Units which eliminated 50 to 60 per cent of applicants, and in spite of Ground Schools, which eliminated 15 per cent of the survivors, there were still at least 6 per cent of men who reached Flying Fields who were discharged or transferred because of inaptitude for flying. This estimate of 6 per cent was obtained from independent data obtained by Thorndike and myself. Thorndike's estimate is based on a follow-up of 1,600 cadets who were admitted to Flying Schools and were either commissioned or discharged from flying status or killed in accident. Observations at Kelly Field showed that on June first, when 750 Reserve Military Aviators had been commissioned, 74 men had been discharged. Of these 55 were discharged for inaptitude or temperamental unfitness for flying, 11 for physical disability or incapacity, 5 for misconduct and 3 for technical reasons. In at least half the cases of discharge for physical unfitness this is but a euphemism for temperamental unfitness for flying. There were about 600 cadets at the Field at the time. Estimating the discharges or transfers that would accrue from this number the estimate of 6 per cent is conservative for Kelly Field. The recent report of the Director of Military Aeronautics for the period ending June 30th, showed that at that time 4,980 men had been commissioned and about 400 had been dropped for inability to fly.

Over and above the 6 per cent who failed completely there are a considerable number who, while ultimately commissioned, have no aptitude for flying, learn with difficulty and smash ships. For example, there were two men who were rated as very poor in flying and who did very poorly in the tests. One took 106 hours for his R. M. A. when the average time was well below 50 hours, and the other took 94 hours. These men were commissioned because it seemed better to use them than to discharge them and train others at great cost in their stead. The 74 men discharged wasted 1,400 hours of flying. Some statistician has figured the total cost of flying training as $2.00 per minute in the air. Whether this includes the cost of wrecked ships is not known, but this group wrecked not less than 30 ships. One man, for example, was given 55 hours of flying instruction and wrecked five ships before he was relieved as "unadaptable for further training." The money cost of the trial and error method of discovering ability to fly, to speak merely of it, becomes enormous. It is unnecessary to multiply evidence to show that the need for improvement in methods for selecting flyers was and is very great.

In June 1917 the National Research Council Committee on Aviation tried out tests reported as promising by French and Italian psychologists, and other tests as well, on 75 candidates at the Massachusetts Institute of Technology Ground School. This work was done by Burtt, Troland and Miles. Stratton, in August 1917, tried out tests on 50 flying cadets at San Diego, and Maxfield tested 44 cadets at Essington Field. Dr. Parsons in the early part of this year tested the cadets in the Naval Aviation Service at Massachusetts Institute of Technology. In this way some forty odd tests had been given something of a try-out. The difficulty of securing satisfactory ratings of flying ability or of an adequate follow-up made the correlations, however, too uncertain for practical conclusions.

In March of this year Captain Stratton and I were detached from Examining Boards to make, in coöperation with Thorndike of the Committee of Classification of Personnel, an official investigation of the most promising of the tests as shown in earlier studies and of others which seemed *a priori* to be worth investigation. It would be too long a story to list the forty tests, give the correlation with flying ability which they showed, and indicate the grounds of selection of tests for further try-out. Ability to fly would seem to require emotional stability and control or "nerve," sensitiveness to tilt, general motor coördination and control, quickness of response to visual and auditory stimuli and to changes of equilibrium, mental alertness, and athletic ability and interest. Tests of most of these traits had shown promise. These, and several others that gave no results worth noting here, were therefore selected and given to 300 cadets and flying instructors at Rockwell Field by Stratton and by myself at Kelly Field. At each field 50 very good flyers, 50 very poor flyers, some of whom had already been relieved from further training for inability to fly, and 50 unknowns were given the tests. Each man was tested on two different days. The tests selected and given at both fields were as follows:

1. *Emotional Stability,* a test which measures the liability of the subject to incoördinate purposeless reaction upon a receipt of a sudden shock, the discharge of a pistol. Measurements were made of amplitude of hand tremor, time of recovery from tremor, changes of frequency of pulse, changes in frequency and amplitude of respiration, and the effect of the shock on rate of solving examples in addition. Of these measurements amplitude of hand response and changes in frequency in respiration were found to be especially significant.

2. *Perception of Tilt,* which measures by means of a

specially constructed tilting chair, the sensitiveness to *gradual* changes of bodily position.

3. *Swaying,* a test of the motor coördination and control required for the subject to stand steadily with the eyes open and with the eyes closed.

4. *Visual Reaction,* quickness of response to movement of the upper hand on the dial of the Hipp chronoscope.

5. *Auditory Reaction,* quickness of response to the noise of the Zimmerman sound hammer.

6. *Equilibrium Reaction,* quickness of response to *sudden* changes of bodily position, to the right and to the left. Burtt's tilting table, which gives a negatively accelerated tilt of constant distance and speed, was used.

7. *Equilibrium Differential,* the sum of the visual and auditory reaction times minus the equilibrium reaction time, a measure of the relative quickness of the response to changes of equilibrium as compared with simple reaction times.

8. *Extension of Curves,* a test roughly simulating the judgments of distance required in landing a ship. This test which showed promise in the earlier study by Stratton gave negative results in the later investigation.

In addition to the psycho-physical tests two tests by means of blank forms were given.

9. *Thorndike's Mental Alertness Test,* an arrangement of thirteen tests of various functions designed to measure native mental ability or general intelligence. This test had been adopted for use by Examining Boards and formed a regular part of examination procedure.

10. *Athletic Achievement and Interest.* Thorndike's study of application blanks filled out by candidates and subsequent performance had shown a positive correlation between athletic ability and success in flying. A more detailed blank was prepared and a scoring system for it worked out by Thorndike and Kelly.

In view of the fact that the Mental Alertness and Athletic Achievement tests were not given to enough men at Rockwell Field, and since they are in the series of tests recommended and adopted, I shall discuss the results from Kelly Field mainly. The instructors and cadets were rated in flying ability by the Officers in Charge of Flying on a basis of judgments of five stage commanders and the progress cards. All officers showed keen interest in the problem and the ratings were made with great care.

The correlations with flying ability, the inter-correlations of each test with every other test, and the intra-correlations

between the trials on two days were determined. Taking into account all available data as well as the later investigations the correlations of the individual tests with flying ability are as follows:

CORRELATIONS WITH FLYING ABILITY

		K. F.
1. Emotional Stability,		
1. Hand Response	.35	.26
2. Changes in Freq. of Respiration	.25	.23
2. Perception of Tilt	.23	.26
3. Swaying	.22	.16
(4. Visual Reaction	0	.15)
5. Auditory Reaction	.14	.15
6. Equilibrium Choice Reaction	.15	.08
7. Equilibrium Differential	.18	—.15
(8. Extension of Curves	—.03	.14)
9. Mental Alertness	.35	.23
10. Athletic Achievement		

The correlations in the first column were obtained by Shepard's formula. Those in the second column are the correlations from the Kelly Field data by the rank method. The figures in parenthesis are for tests not in the recommended series.

The best tests, as the table shows, are emotional stability, especially the hand response, perception of tilt, and mental alertness. "With the material used the correlation between the Thorndike-Kelly athletic score in flying ability was over .6 though certain factors were operative which make this possibly too high and the number of cases was under 100." None of these tests alone obviously is a sure symptom for inaptitude for flying or a sufficient basis for prognosis. When, however, the individual measurements are combined the composite score obtained from the series or team of tests recommended gives a multiple correlation coefficient of .70, which is a sufficient basis for prognosis with some confidence. In selecting the team of tests those were chosen which show the highest correlations with flying ability, satisfactory reliability coefficients, and low correlations with each other. To determine the weights to be given each test toward the composite score would ideally require the calculation of partial correlations and coefficients of regression. The time and labor for this analysis has not been available. A provisional scoring and weighting scheme, scores being expressed as deviations above and below the mean of performance in the tests, was worked out. The weights were estimates made with due regard to the principles and factors that determine weighting by the partial correlation method. The success with which the recommended

tests segregate the poors and the goods is indicated in the following table:

SCORES ON FLYING ABILITY OF 100 CADETS AND INSTRUCTORS AT KELLY FIELD

Score in the recommended series of tests	Report from flying school p-poor. g-good.	Notes to date of June 15	Average dual time of others
—75 or worse	pppppp	4 suspended, relieved or discharged	15:50
—50 to —74	ppppppg	4 " " " "	11:16
—25 to —49	ppppppp pppppppg	5 " " " "	10:17
0 to —25	ppppggg ppppggg ppppggg pppp	9 " " " "	10:03
0 to —25	ppggggg	1 " " " "	9:19
—25 to —49	pggggggg pggggggg	1 not reported	9:23
—50 to —74	ggggggg		7:50
—75 or better	g		7:21

The acid test is, of course, the ability to predict. In order to eliminate the possibility of confusing abilities produced by flying with those prophetic of ability to fly, the tests were given to 50 cadets of unknown ability who had just reached the Field. Some of these had had no experience in flying and none over two or three hours of dual instruction. On a basis of performance in the tests solely I submitted to the Officer in Charge of Flying a prediction that five of these men would either be discharged or learn to fly with great difficulty. Two others were selected as showing special aptitude, one of whom because of deficiency in formal education had difficulty in getting into the Service. Of the five men three were discharged after 4, 20 and 22 hours of instruction respectively. One of these men wrecked two ships completely. The fourth man was relieved from flying at one time and brought before the Discharge Board but was given another chance and finally was commissioned after 85 hours of flying instruction when the median time for his class was 60 hours. The fifth man was commissioned after 93 hours of instruction. The Officer in Charge of Flying said of these men that they did "Fair" work. The two men who were picked to show exceptional aptitude were commissioned as pursuit pilots and their work characterized as "Very Good." One of the men made remarkable progress and was commissioned after 43 hours of flying when the median time was 70 hours. This was the man who barely met the educational standards.

The success with which the tests differentiated the goods from the poors and the correspondence between prediction and performance were sufficiently convincing so that the Personnel Section of the Air Service was authorized to proceed with their introduction. Steps were taken to organize four Personnel Units to be established at the four Ground Schools and in affiliation with four Examining Boards. The all too familiar vexatious delays in securing necessary personnel and equipment prevented their being fully organized and actually in operation before the armistice was signed and Examining Boards and Ground Schools were closed.

While waiting for personnel and equipment at Princeton, where the training unit was located, Captain F. L. Wells and I carried out a series of experiments with four subjects to determine (1) the susceptibility of the tests to systematic practice, (2) the advisability of repeating the tests on the same day instead of as heretofore on the next or some subsequent day, and (3) the effect of time of day on performance in the tests. In the practical administration of the tests questions on these points would naturally arise and did arise. The tests, except those of mental alertness and athletic achievement, measure relatively simple functions that enter into flying ability. *A priori* one might expect, though one could not be sure without actual experiment, that these tests, having been highly practiced by the ordinary conditions of life, would not show very quick or very marked effects of practice or of diurnal variation. If this supposition holds, the tests are for practical purposes all the more valuable. The results show

1. That except for the equilibrium choice reaction practice effects are slight for daily tests over a period of fifteen days. The practice effects, moreover, are least evident in those with the poorest initial scores.

2. That the variability in the once-a-day tests is but little greater than in the twice-a-day tests.

3. That time of day for all practicable purposes is negligible.

The Air Service is certain to become an increasingly significant arm of the military forces and it is equally certain that aviation will have ere long important applications in civil life. The selection of the apt and the elimination of the unfit for flying is, therefore, so important that the responsible authorities should on a basis of the showing made be required to give the tests a practical tryout and take steps for their improvement, in order that vast sums should not continue to be expended fruitlessly and human life wasted.

MENTAL TESTS FOR PROSPECTIVE TELEGRAPHERS, A STUDY OF THE DIAGNOSTIC VALUE OF MENTAL TESTS FOR PREDICTING ABILITY TO LEARN TELEGRAPHY

By L. L. Thurstone, Carnegie Institute of Technology, Pittsburgh

This investigation was made to ascertain whether it is possible to predict ability in telegraphy by mental tests. Our subjects were 165 drafted men of Class I, who registered for a night course in radio-telegraphy. The course was given at the Carnegie Institute of Technology to prepare drafted men for the Signal Corps of the U. S. Army. The course opened on November 19, 1917, and this report is written in March, 1918.

The two main features in this investigation are the mental tests and the daily records of progress. The mental tests were given to the men immediately after registration for the course on the same evening that they were given the physical examination. Every applicant took both the mental and physical examination before entering classes. The records of progress were obtained by means of prescribed tests of receiving speed. The instructions and copy for the tests of speed were prepared in the office of the Psychology Department and administered to all of the fourteen classes with as much uniformity as possible. The receiving speed test was of two minutes duration, given at every meeting of the class and always between the first and second hours of the three-hour evening session. A more complete description of these tests and the directions for administering them are contained in my "*Handbook for Teachers of the Telegraphic Code*," which has been published as Bulletin No. 16 of the Federal Board for Vocational Education, Washington, D. C.

I. SUBJECTS

1. *Occupations.* Among the 165 students were represented the mechanical and electrical trades, building and steel trades, engineers, chemists, salesmen, bookkeepers, etc. It is of interest to note that there is no occupational differentiation in the ability to learn telegraphy. Hence, occupation is not suitable as an index for selecting prospective telegraphers.

2. *Schooling.* Here, again, we find the rather surprising fact that schooling is not suitable as an index of ability to learn telegraphy as far as this type of group is concerned. It is quite likely that the factor of schooling is loaded in this group of students so that the brightest of the high school and college graduates may have found preferred service in the army before being drafted. If this is true, then the college, high school, and grammar school groups here represented are not comparable. At any rate, if we take our findings at face value, *ability in telegraphy cannot be predicted on the basis of general schooling.*

The Pearson correlation coefficient between number of years of schooling and receiving speed in telegraphy after one hundred hours of practice is 0.00 ± 0.08; ($n=158$). While the correlation coefficient is low, it may be significant that the eleven students who left school before finishing the eight grades were conspicuously low in telegraphy.

3. *Age.* Being a group of drafted men the age limits were 21 to 31 years. The frequencies for the higher ages within the draft limits are relatively low. This is explained by the greater frequency of married men at those ages and by other causes of exemption among the older of the men of draft age. It is not without interest to note that the older men exhibited more perseverance in staying by the rather monotonous evening buzzer practice than the younger men.

The correlation between age and receiving speed in telegraphy after three months practice (about 100 hours) is -0.09 ± 0.08, which indicates that age has no diagnostic value in predicting ability in telegraphy within the age limits of 21 and 31.

II. THE MENTAL TESTS

All of the students who entered the radio course during November and December took a psychological examination consisting of eight mental tests. The purpose of these tests was to ascertain whether mental tests could be used to predict ability to learn telegraphy.

We gave the following tests:

1) Rhythm Test
2) Opposites Test
3) Analogies Test
4) Gordon Directions Test
5) Trabue Completion Test
6) Spelling Test
7) Arithmetic Test
8) Sentence Test

All of these mental tests have been checked up with the highest receiving speed attained during the first one hundred hours of practice.

1. *Rhythm Test.* This test was designed specially for the purpose of testing ability in telegraphy. It is given as follows: The examiner is equipped with a buzzer and telegraphic key. The examiner sounds on the buzzer a rhythm pattern consisting of auditory dots and dashes. Each pattern is repeated once in order to give all students a fair chance to hear it. They then reproduce the pattern, using a short horizontal line for a dash and a short vertical line for a dot. The test consists of thirty-five of these rhythm patterns given in the order of difficulty, beginning with the easiest. The score in the rhythm test is the number of errors made. The rate of presentation is such as would correspond to a telegraphic sending speed of ten words per minute and the whole test is given in about fifteen minutes.

The rhythm test gives a correlation of $+0.48 \pm 0.06$. Of the highest thirteen students in this test, only one did worse than the average in telegraphy. *Of the lowest fifteen students in this test, only three students attained the average of the class.*

2. *Opposites Test.* This is a good test for predicting ability in telegraphy. The correlation coefficient for speed in receiving telegraphic code and scores in the Opposites Test is $+.42 \pm 0.04$.

The ten radio students who rank lowest in the Opposites Test are all inferior in telegraphy. Not one of them attained even average performance in telegraphy. Their average receiving speed is 3.8 words per minute as compared with 6.4 words per minute for the whole group.

The lowest twenty per cent of the group (23 students), arranged according to the Opposites Test, have an average receiving speed of 4.5 words per minute as compared with 6.4 for the whole group. Of these 23 students, only five exceeded the average of the group in telegraphy.

3. *Analogies Test.* The Pearson correlation coefficient for receiving speed in telegraphy and the Analogies Test is $+0.29 \pm 0.05$.

Analyzing the scatter diagram more closely, we find that *of the seventeen students who constitute the lowest ten per cent of the group, according to the Analogies Test, only one student attained the average of the class in telegraphy.* The average speed in telegraphy for these seventeen students is 5.2 after one hundred hours of practice. The average speed in teleg-

raphy for the upper ten per cent of the group, divided according to the Analogies Test, is 7.5 words per minute after one hundred hours of practice. The average speed in telegraphy for the whole group at the same stage of practice was 6.4 words per minute.

This indicates clearly that while the Analogies Test does not give a high correlation coefficient it can be used to eliminate a considerable portion of those who lack the special ability to learn telegraphy. The average telegraphic speed for the lowest half of the group, divided according to the Analogies Test, is 5.4 words per minute, the speed for the upper half of the group is 7.1 words per minute.

If the lowest twenty per cent of the group had been eliminated on the basis of the Analogies Test the average speed in telegraphy would have been 7.0 words per minute instead of 6.4 words per minute.

4. *Gordon Directions Test.* The correlation of this test with speed in telegraphy after one hundred hours of practice is $+0.27 \pm 0.05$. It has diagnostic value for ability in telegraphy for the lowest ten per cent of the group arranged according to the test, but not for the middle and higher scores in the test. *Of the 33 radio students who scored in the lowest twenty per cent of the group, only four succeeded in attaining the average of the class in telegraphy.* The average receiving speed for this lowest twenty per cent of the group, divided according to the test, was 4.6 words per minute, the average for the whole group was 6.4 words per minute. If those registrants who scored in the lowest twenty per cent of the group had been eliminated at the start the average receiving speed for the class would have been raised from 6.4 words per minute to 6.9 words per minute.

If those radio students who were in the lowest half of the group according to the Directions Test had been eliminated the average receiving speed of the class after 100 hours of instructions would have been raised from 6.4 words per minute to 8.1 words per minute.

The two most important facts about the use of the Directions Test for selecting telegraphers are as follows:

(1) *If we consider all those students as failures who did not exceed three words per minute in receiving tests after 100 hours of practice, the test would enable us to exclude 48 per cent of these failures;* (2) only 12 per cent of those who score below 40 in the test succeed in doing better than the average of the class.

5. *Trabue Completion Test.* We used the Bureau of Salesmanship form of this test. It gives a correlation with telegraphy of $+0.21 \pm 0.05$. Of the thirteen lowest students in this test, four students attained the average of the class in telegraphy. Of the eleven highest students in this test, only two were below the average in telegraphy. While this test is not without some diagnostic value for ability in telegraphy its best use is as a supplement to other more successful tests. This test is usually a good index of general intelligence, but in the present investigation our criterion is ability in telegraphy which is a special ability and that probably accounts for the rather low rank of the completion test.

6. *Spelling Test.* This test was given to the radio students in the belief that ability in spelling might prove to be of diagnostic value in the selection of telegraphers. The correlation between speed in telegraphy and score in the spelling test is $+0.18 \pm 0.06$.

Although this coefficient is lower than one would expect, the spelling test can be used effectively for eliminating some failures in telegraphy. *Of the 12 students who made the lowest scores in spelling, only two attained the average of the class in telegraphy.* The average speed in telegraphy for these twelve students was 4.8 words per minute as compared with 6.4 words per minute for the whole group.

7. *Arithmetic Test.* The correlation of this test with speed in telegraphy is only $+0.08 \pm 0.05$. Inspection of the scatter diagram shows that this test has no diagnostic value for ability in telegraphy.

8. *Sentence Test.* This test gives a correlation with telegraphy of $+0.09 \pm 0.05$, which indicates that the test has no diagnostic value for ability in telegraphy.

III. SUMMARY

(1) Figure 1 shows in black the speed in telegraphy of those who failed in any two or more of seven tests. The seven tests here considered are: Rhythm, Opposites, Analogies, Directions, Sentence Completion, Spelling, Number Completion. The white spaces indicate the telegraph speed of those who did not fail in more than one test.

(2) Figure 2 shows the scatter diagram for the Rhythm Test with a Pearson correlation coefficient of $+0.48 \pm 0.06$.

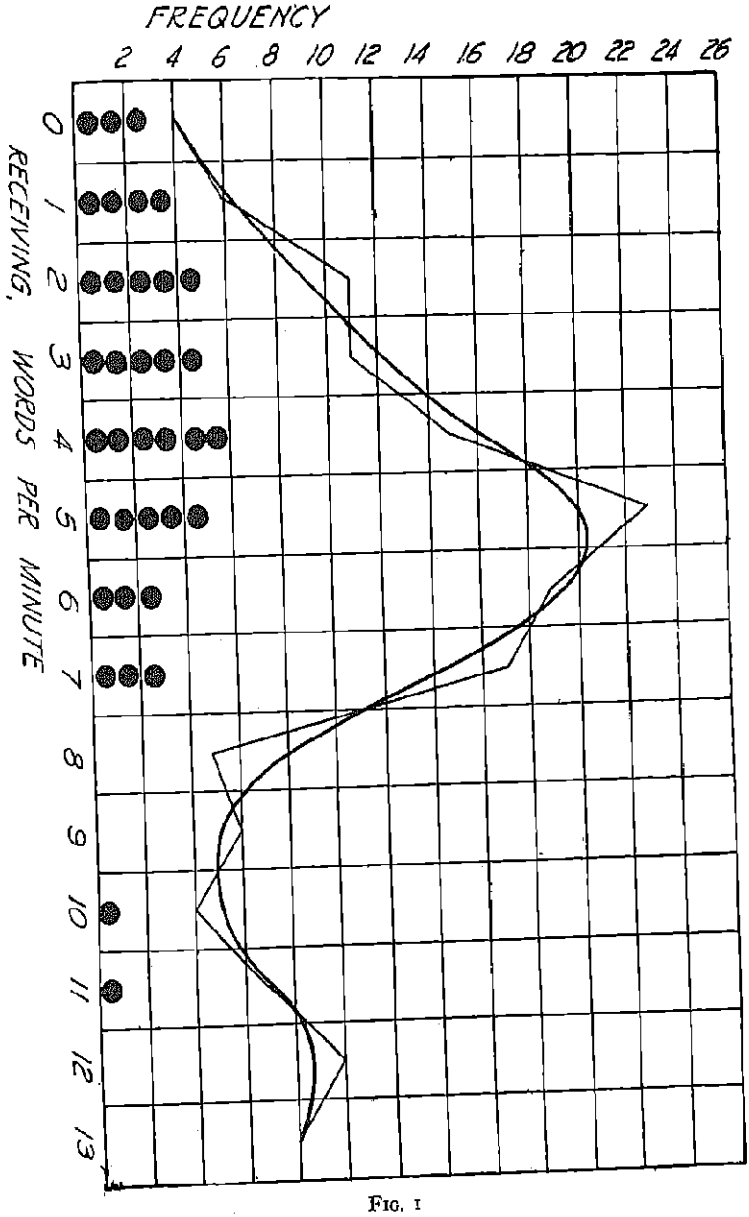

FREQUENCY
2 4 6 8 10 12 14 16 18 20 22 24 26
RECEIVING, WORDS PER MINUTE
0 1 2 3 4 5 6 7 8 9 10 11 12 13

FIG. 1

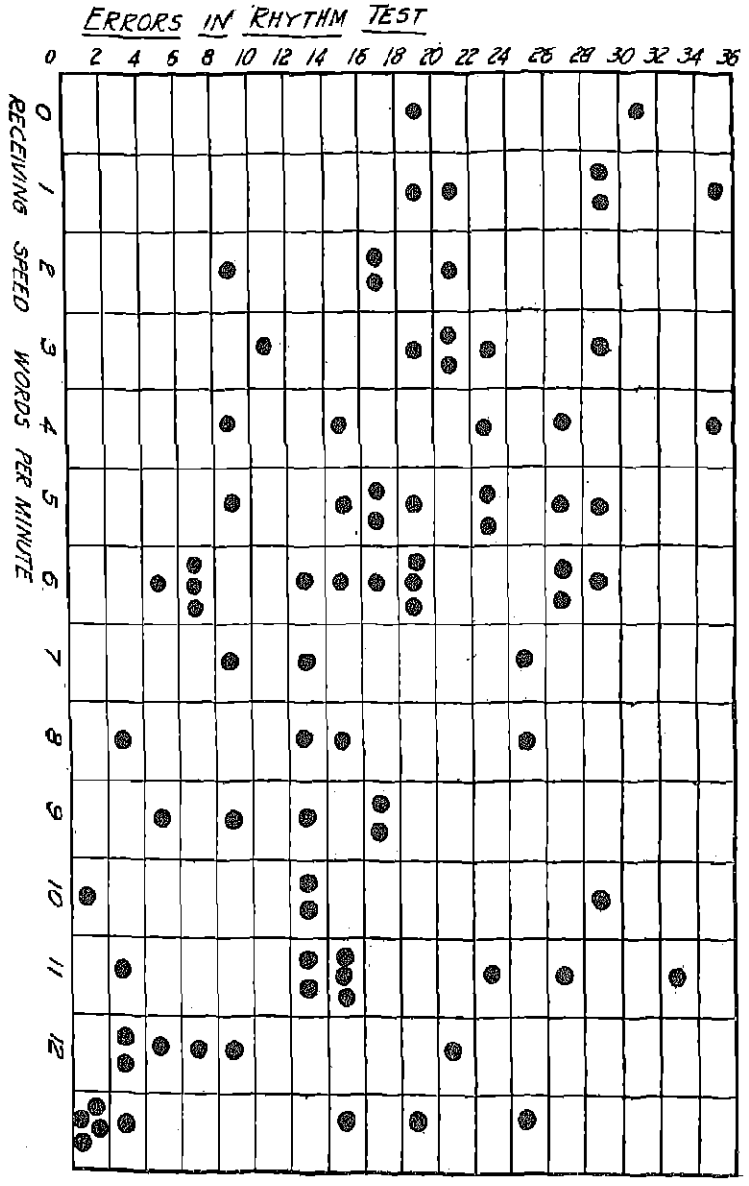

ERRORS IN RHYTHM TEST
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
RECEIVING SPEED WORDS PER MINUTE
0 1 2 3 4 5 6 7 8 9 10 11 12

FIG. 2

(3) Five of the tests have been evaluated by the method of multiple correlation: These tests were: Rhythm, Opposites, Completion, Analogies, and Gordon Directions. The fourth order coefficients for these tests with telegraphic receiving speed after one hundred hours of practice as the criterion are as follows:

Rhythm: $r_{12\ (3450)} = +.31$
Opposites: $r_{13\ (2450)} = +.21$
Completion: $r_{14\ (2350)} = +.096$
Analogies: $r_{15\ (2340)} = +.02$
Gordon Directions: $r_{10\ (2345)} = +.06$

(4) The general intelligence tests are not as valuable for diagnosing ability to learn telegraphy as for measuring general intelligence. Ability in telegraphy is probably a special ability.

(5) The fact that years of schooling does not agree with ability to learn telegraphy indicates that this is a special ability. College graduates usually do better on general intelligence tests than those who have only finished grammar school. But college graduates do not necessarily excel in learning telegraphy.

(6) The total correlation coefficient $R_{1(23456)}$ for receiving speed in telegraphy on the one hand and the combined score in five tests is $+.53 \pm 0.04$. The Rhythm Test alone gives a zero order coefficient of .48 which is not noticeably raised by the addition of four of our best general intelligence tests. Hence the Rhythm Test carries most of the load.

(7) The bimodality of figure 1 is probably significant and indicates that we are here dealing with a special ability. The upper mode probably represents that part of the group of subjects in which the special ability is present.

(8) A vocational survey of students in a radio mechanics course shows that schooling is one of the best diagnostic criteria in selecting men to be trained in the care and repair of wireless apparatus but our analysis also shows that when selecting men to be trained as telegraph operators the schooling is of no diagnostic value.

AN ABSOLUTE INTELLIGENCE SCALE: A STUDY IN METHOD

By Grace Arthur and Herbert Woodrow, University of Minnesota

Aim: The present investigation[1] is a study of a series of group tests as regards their value,—

(1) in classifying children of grade school age according to mental ability;

(2) in the construction of a scale of absolute intelligence units;

(3) in comparing the "more or less" with the "all or none" method of scoring; and

(4) in a system of weighting, which is neither arbitrary nor determined by subjective evaluation of the tests used, but which is logical in its origin and automatic in its application.

The aims of this investigation need but little explanation. The need for group intelligence tests has long been recognized and several systems of such have made their appearance since the present investigation was undertaken. Our interest, however, is not primarily in the development of group tests, but in certain fundamental features of method, of the same importance in individual testing as in group testing. The feature in which we are mainly interested is the use of absolute intelligence units,—units, that is, which are equal at all points of the scale and in terms of which it is possible to plot an absolute intelligence curve showing (in so far as the tests used really measure intelligence) the absolute increase in intelligence of children with increase in chronological age.

In the selection of tests, we have chosen only such as permit a measurement of degree in the ability tested, whatever that degree of ability may be. Such tests offer a marked contrast with the Binet-Simon tests. Many of the latter determine merely whether or not a certain degree of ability is exceeded, instead of measuring by how much it is exceeded or by how

[1] It is with pleasure that acknowledgment is made to Dr. Frances Lowell, of the University of Minnesota for suggestions as to the choice of tests and details of procedure, and to Miss Edith Taylor, principal of the Hancock School (the school tested), St. Paul, Minn., for her aid in securing a favorable attitude towards the tests on the part of teachers and pupils.

much it fails of being reached. Such tests, the child either fails or passes. They may be termed "all or none" tests. In contradistinction to them, the tests here employed may be called "more or less" tests, because they measure the ability which they test in any amount from zero up to the maximum.

When scored by the more or less method, different tests, such as must be used in any measurement of intelligence, give measures in different units. It becomes an important problem, consequently, how to average together the results of the different tests. A common method of combining the results of several tests is to weight them, but the methods of weighting heretofore used appear unsatisfactory,—and even, in some conspicuous cases, entirely arbitrary. We have found, however, that by using absolute intelligence units of the sort mentioned, weighting of the different tests occurs logically and automatically, and in exact correspondence to their value in discriminating between children of different age and, therefore, presumably of different intelligence.[2]

Choice of Tests: As the tests were to be given to all the pupils of a grade school, it was necessary to select such as would be simple enough for the younger children to comprehend, and at the same time difficult enough to tax the ability of the most mature. The tests finally decided upon were the following:

Immediate Memory Span (digits)
Easy Opposites (new lists based on the data obtained by King and Gold)[3]
Hard Opposites
Substitution (Woodworth—Wells)
Word Building (*a, e, i, r, l, p,* and *a, e, o, b, m, t*)
Language Completion (Trabue, Scales B and L)
Anagrams (original)
Cancellation (three letter)
Comprehension (adapted from Kuhlmann)

The only ones of these tests requiring special description are the anagrams test and the comprehension test. The anagrams test consisted of twenty-one words, varying in length from three to seven letters, taken from the lower end of the Ayres' spelling scale. A definite scheme of letter displace-

[2] It hardly needs to be pointed out here that the only natural intelligence scale by which the validity of psychological scales can be checked is that formed by the chronological age groups.

[3] The Standardization of Certain Opposites Tests. *Journal of Educational Psychology,* 1916, p. 459.

ment was followed and the words were arranged in order of difficulty estimated from their relative length and the number of letters displaced. This trial list was given to the pupils of the first, second, sixth and eighth grades of another school, and the words were then arranged in increasing order of difficulty as shown by a frequency table based upon the responses of these pupils. Spelling ability, as a factor of success in this test, was reduced to a minimum by using only the easiest words from the Ayres' list. To offset the puzzle element as much as possible, the subjects were instructed, in case they came to a word they could not get, to go on to the next and to come back to the hard word later.

For the comprehension test, three of the best individual tests from the Kuhlmann 1917 revision[4] were combined and adapted for group work. Two test sheets were used, the easier for all grades, the harder for the 4th to the 8th grades inclusive. Where results from both forms were used, they were added to give the total score.

Subjects: The subjects were grade school pupils from 6 to 17 years of age, inclusive. Norms for the separate tests were secured for the ages 6 to 13 years (though the system of tests here employed is not regarded as adapted to children under 7 years). Results for the age groups above 13 were discarded, because all the members of these groups were retarded pedagogically. Eleven subjects of the 13 year group were one-half to one year advanced pedagogically, and as they entered high school before the testing was completed, they had to be rejected from the cases finally used. From an examination of the age and grade distribution of the school, it would appear that the 10 year group had fewer bright subjects than had the other age groups. It is probable, therefore, that the 10 years norms, as well as the 13 year norms, are too low. This is of little concern to us in the present stage of our investigation,[5] as we believe we can improve on the selection of tests used and expect to obtain new norms.

For all subjects, chronological ages were reckoned to March 8th, 1918, as that date would give an error of less than one month in the great majority of the cases.

Instructions: Written instructions were used by the experimenter for all the tests, and were half spoken, half read. For the most part, the instructions were adaptations of those of

[4] Kuhlmann, The Measurement of Mental Development. *Journ. of Psycho-Asthenics, Monog. Sup.*, 1917.

[5] This work was interrupted by the entrance of one of the writers into military service, but will be continued by the other.

earlier investigators. They were given a preliminary trial with first, second, sixth and eighth grade groups of another school, and were then revised on the basis of the results with these groups.

In giving instructions, group response in a trial test was secured wherever possible. Purity of diction was sacrificed in every case where a colloquial form appeared to be more readily understood. We spare the reader details of the tests and instructions, hoping to present them at a future time when our scale of mental measurement is more perfectly standardized.

Norms: With each age group, the mean, as the most probable true value, was taken as the norm. The mean was found to increase from one age to the next for all tests except the cancellation test, for which the 9 and 10 year scores are identical. It is impossible to make any direct comparison between our norms and those of other investigators, on account of differences in the form of the tests or the method of giving them.

The norms obtained for each of the nine tests is given in Table I, in the lines marked *Av.* Immediately below each norm, that is, below the average score of each age group, is given the corresponding standard deviation (σ). Our purposes would probably be served by any other measure of variability about as well as by the standard deviation. At the top of Table I, are given the ages of the children and the number (No.) at each age. By age 6.5 is meant all ages from 6 yrs. 0 ms. to 6 yrs. 11 ms. inclusive; similarly for all other ages.

Point Scale Norms for Individual Tests: To deal with the data effectively, it is of course necessary to reduce the norms found for the different tests to comparable units. These units should not only be comparable, but should also be of such a nature that the test which has the highest discriminative value will have its norms represented by the largest number of units, so that, in summing up the total number of units, or points, scored by a child in all tests, the different tests will have weight directly in proportion to their discriminative value.

The discriminative value of a test consists in its capacity for bringing out individual differences in intelligence. This value is low when the scores of children of one age group largely overlap those of adjacent age groups; and it is high when there is little overlapping. Now the amount of overlapping in the scores of two adjacent age groups depends upon

two factors: first, the difference in the average score of the two groups, and second, the variability within those groups.

That the amount of overlapping will vary with difference in the average scores may be readily appreciated. If there is no difference between the average scores of two groups, their scores will substantially coincide, in fact exactly so, except for difference in the form of distribution of the scores. In such a case, if 50 per cent of one of the groups exceeds its own

TABLE I.
NORMS

Tests.	Age No.	6.5 66	7.5 87	8.5 82	9.5 83	10.5 98	11.5 95	12.5 87	13.5 71
Easy Opposites	Av.	0.0	1.1	2.6	4.4	5.2	6.5	7.4	8.6
	σ	*0.0*	*1.2*	*1.5*	*2.5*	*2.1*	*1.9*	*2.1*	*2.6*
Hard Opposites	Av.	0.0	0.5	1.3	2.8	3.7	6.3	7.1	9.1
	σ	*0.0*	*1.0*	*1.5*	*3.0*	*3.1*	*4.4*	*4.1*	*4.5*
Substitution	Av.	10.1	20.5	30.2	40.3	44.2	50.6	52.3	59.1
	σ	*6.2*	*11.1*	*11.5*	*12.0*	*12.5*	*11.2*	*10.7*	*13.1*
Word Building	Av.	0.4	2.3	4.3	6.3	7.4	9.8	10.3	12.0
	σ	*0.9*	*2.5*	*2.8*	*3.4*	*3.2*	*3.9*	*3.7*	*3.0*
Digits' Span.	Av.	3.0	4.4	4.9	5.3	5.4	5.7	5.9	6.0
	σ	*1.5*	*0.8*	*0.7*	*0.8*	*0.8*	*0.8*	*0.8*	*1.0*
Language Completion	Av.	1.7	5.0	7.5	10.3	11.4	14.2	15.0	15.1
	σ	*2.2*	*3.0*	*3.1*	*4.4*	*3.6*	*4.7*	*3.8*	*4.4*
Anagrams	Av.	1.8	4.5	7.4	9.1	9.2	10.2	11.1	11.7
	σ	*2.2*	*3.5*	*3.1*	*3.1*	*3.2*	*2.8*	*2.8*	*3.4*
Cancellation	Av.	12.1	16.0	17.2	20.6	20.6	23.0	23.7	25.6
	σ	*5.2*	*4.7*	*3.9*	*5.0*	*4.9*	*5.1*	*5.7*	*5.6*
Comprehension	Av.	2.5	4.0	5.8	9.5	13.1	16.1	16.8	17.2
	σ	*2.5*	*2.0*	*3.9*	*6.2*	*5.3*	*3.9*	*3.3*	*2.9*

average, we may reasonably expect that substantially 50 per cent of the other will also exceed that average. On the other hand, if there is a great difference between the average scores of the two adjacent groups, there will be little overlapping, and if 50 per cent of both groups exceed their group average, a very high percentage, well over 50, of the older group will exceed the average of the younger group. Obviously, if the variability is the same, the greater the difference in the percentage of the two groups exceeding a fixed degree of ability, the less the two groups will overlap. In general, with a fixed

variability, the amount of overlapping varies with the difference between the averages of the two groups. This difference between adjacent averages, then, is one factor determining the *discriminative value* of the test.

Equally important is the factor of variability. With a fixed difference between the two averages, the amount of overlapping between the two groups depends upon their variability. We may have, for example, a difference of five units between the two averages, and have either no overlapping or a high degree of overlapping, depending upon whether the children of each group are distributed closely about their own average or spread out over a wide range. If the variability is small, we may find that 100 per cent of the older group exceeds the median performance of the younger; if large, that less than 60 per cent of the older exceeds the median of the younger. As a measure of the variability of any two adjoining age groups, we have used the average of the standard deviations of those two groups.

To find the discriminative value of any test, then, between two adjoining age groups, it is necessary to compare the difference between the average scores of the two groups with the average of their standard deviations. The larger the ratio of the former to the latter, the greater will be the discriminative value. Accordingly we may write, as the formula for calculating discriminative value,

$$\text{D. V.} = \frac{\text{Av}_2 - \text{Av}_1}{\frac{1}{2}(\sigma_1 + \sigma_2)}$$

in which *D. V.* stands for discriminative value, Av_2 and σ_2, respectively for the average and standard deviation of the older group and Av_1 and σ_1, for the same of the younger group.

While the above formula is the one we have used in calculating the discriminative values of our tests, it may aid in understanding their nature to point out that they may be calculated in other ways. This is possible because of the relationship between the standard deviation and the percentage of children exceeding any given deviation from the average. On account of this relationship, discriminative value may be calculated from the percentage of children of one age who pass the average score of those of the next higher age; or from the percentage of the higher age failing to pass the average of the next lower age; or, lastly, from the average of these two percentages. Still another way to put the matter, is to say that discriminative value depends upon the increase with age in the percentage passing a given score.

One way, then, to calculate the discriminative value of a test between two adjacent age groups would be on the basis of the average of two percentages, one, that of the lower group exceeding the average of the higher, the other, that of the higher failing to reach the average of the lower. If the average of these two percentages happened, for example, to be 15.86, then we could conclude that the difference between the two group averages just equalled the average of the standard deviations. This would mean, according to our formula, that the discriminative value of the test in question was 1.0. If the average of the two percentages was anything else than 15.86, we would simply look up the fraction of the standard deviation to which it corresponds. For example, if it was 10, the value of the test in discriminating between the two groups would be 1.28; if it was 35, the discriminative value would be only 0.39.[6]

Again, if we have as data the fact that 75 per cent of the higher group exceed the average for the lower group, then we can, by the use of published tables,[7] read off directly the value of the test in discriminating between the two groups. In the instance given, the discriminative value would be 0.67.

Now it is well known that the percentage of one age group passing the norms of the age group one year younger tends in general to decrease as we reach the higher chronological ages. We should expect, then, that the discriminative values of our tests would show a tendency to decrease at the higher chronological ages. As a matter of fact, they do so.

Before leaving the subject of the relation between discriminative value and the percentage of one age group passing the norms of the preceding age group, it is of interest to consider one conclusion generally conceded to follow from the variation in this percentage. At the higher chronological ages, as already stated, the percentage of any year group passing the norms of the one year younger group is usually less than at the lower chronological ages. This fact, it is commonly recognized,[8] means that the absolute difference between the mental ages of two groups differing by one year is less at the upper chronological ages than at the lower,—that a year of mental age is a smaller unit at the upper ages than at the

[6] See Thorndike, *Mental and Social Measurements*, 1904, p. 148.

[7] Thorndike, *loc. cit.*, or W. F. Sheppard, *Biometrika*, 1903, Vol. II, p. 182.

[8] *Kuhlmann*, Some Results of Examining a Thousand Public School Children with a Revision of the Binet-Simon Tests of Intelligence by Untrained Examiners, *Journ. of Psycho-Asthenics*, Vol. XVIII, 1914, p. 235.

lower. As Kuhlmann puts it, "the amount of mental development during a year cannot be taken as an accurate unit of measurement of intelligence."[9] Mental age is not an absolute unit, but a changing unit. The present method subscribes to this point of view. In fact, it might be summarized as an attempt to find, for the different ages, norms which correspond in magnitude to the actual absolute increase in intelligence with age.

In the present work, the discriminative value of a test as derived from the formula $Av_2 - Av_1 / \frac{1}{2} (\sigma_2 + \sigma_1)$, has been used in preference to that based upon the obtained per cents of one age group falling above or below the norms of adjacent groups. The discriminative values when based upon the standard deviations, as in the formula we have used, are based upon the data for the entire group instead of for portions lying towards the extremes. We have worked through all our data, however, by both methods, and found that in the great majority of instances the results derived by the two methods are so nearly identical as to make it feasible to use whichever is the more convenient. In further work, therefore, while norms are shifting, in order to avoid the labor of calculating the standard deviation anew every time new cases are added, it will probably be advisable to use discriminative values based upon the per cents of the various age groups passing the average of the next higher group and failing to pass the average of the next lower group.

The discriminative values may be treated as "points" or units. Thus, if the discriminative value of a given test between the ages 6.5 and 7.5 is 1.4 units, we consider that the 7.5 year group shows an increase of 1.4 units over the 6.5 year group in the ability tested. We do not, of course, know how many units of ability the 6.5 year group possesses, as we did not extend our testing down to the age of zero, whatever that may be. Taking the average ability of the 6.5 year group as x, however, we may say that the 7.5 year group shows $x + 1.4$ units of ability. Similarly, the quotient representing the value of the test in discriminating between the 7.5 year group and the 8.5 year group indicates the value to be added to the 7.5 year average of ability. Thus, if the discriminative value of the test for the age groups 7.5 and 8.5 is 1.1 units, then we can represent the average 7.5 year ability as $x + 1.4$ units, and the average 8.5 year ability as $x + 1.4 + 1.1$, or $x + 2.5$ units. By continuing this process through all age groups, the original norms are readily changed to norms in terms of points.

[9] *Loc cit.*

In the case of tests of high discriminative value, more points are added for each year, and the norms thus come to be expressed as a larger number of units than in the case of tests of low discriminative value. Because of the different rate of increase in points or units from age to age in the different tests, it results that when we add together the scores for all tests to obtain the final scale, there is an automatic weighting in favor of the more valuable tests. This may be clearly seen, by a comparison of the point scale for easy opposites with that for cancellation. The subject who just equals the 13.5 year norm in easy opposites will secure $x + 4.9$ points, while the subject who equals the 13.5 year norm in the three letter cancellation test will earn only $x + 2.8$ points. This weighting is not done arbitrarily or by any uncertain subjective estimate; nor is it done simply because the opposites test is regarded as a good test and the cancellation a bad one; it is done because the former, by exhibiting at all ages a higher degree of discriminative capacity than the latter, has secured at each age a larger number of points for its norms. How this comes about may be seen from the following table (No. II).

TABLE II.

COMPARISON OF POINT NORMS OF EASY OPPOSITES WITH THREE LETTER CANCELLATION

	EASY OPPOSITES				CANCELLATION			
Age	Original Norms	σ	Discrim. Value	Point Norms	Original Norms	σ	Discrim. Value	Point Norms
6.5	0	0	0.9	x	12.1	5.2	0.8	x
7.5	1.1	1.2		x+0.9	16.0	4.7		x+0.8
8.5	2.6	1.5	1.1	x+2.0	17.2	3.9	0.3	x+1.1
9.5	4.4	2.5	0.9	x+2.9	20.6	5.0	0.8	x+1.9
10.5	5.2	2.1	0.34	x+3.2	20.6	4.9	0.0	x+1.9
11.5	6.5	1.9	0.65	x+3.9	23.0	5.1	0.5	x+2.4
12.5	7.4	2.1	0.40	x+4.3	23.7	5.7	0.1	x+2.5
13.5	8.6	2.6	0.55	x+4.9	25.6	5.6	0.3	x+2.8

The increase with age in the point values of the easy opposites and cancellation norms is shown graphically in Figure No. 1. This figure shows that growth in the ability tested by the easy opposites test is very much greater, when measured in absolute units, than is that in the ability tested by the cancellation test, this in spite of the fact that the original score norms show just the reverse. The increase between seven and thirteen years in the number of cancellations is 13.5, whereas the increase in the number of easy opposites is only 8.6; but the absolute increase in ability with age is nevertheless much greater in the case of the easy opposites.

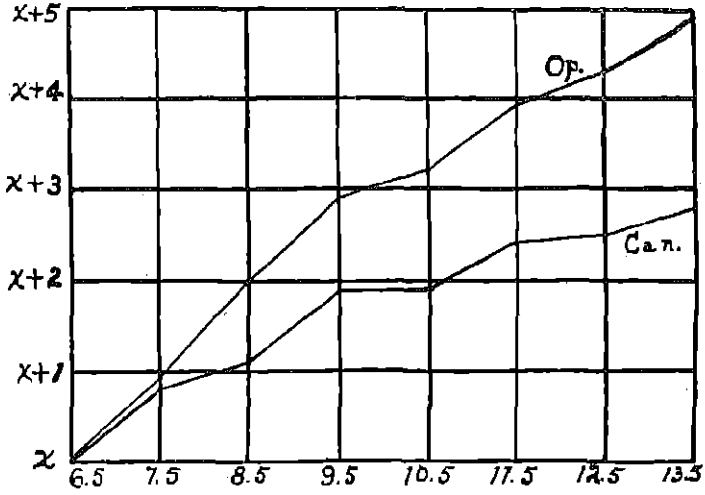


FIGURE 1.— Point norms for two tests, showing weighting effect. Op., opposites test; Can., cancellation test. Chronological ages represented by abscissæ, points by ordinates.

It might appear, at first thought, that in combining the point norms of the individual tests to obtain the final scale, the different tests are not weighted to the same extent toward the x end of the scale as towards the upper end. For example, in Figure 1, since the two curves start at the same point, it may be asked, are they not weighted equally at this point? Such a question, however, would show a misconception of wherein the weighting consists.

To understand this matter, it is necessary carefully to analyze our procedure. What we do, in reality, is simply to combine (by adding their ordinates) a number of curves like those

in Figure 1 to obtain one final curve. Now the *shape* of this final curve is not affected by the height at which we place the individual curves. For example, if we combine the two curves of Figure 1 (by adding their ordinates), the shape of the resulting curve is the same whether the individual curves are plotted so as to coincide at the age of 6.5, as in the figure, or, on the other hand, so as to coincide at the age of 13.5. If we plotted all our individual curves so that they would coincide at the age of 13.5, then x would be the number of units of intelligence possessed by the average child of 13.5, and we would use the scale to measure the intelligence of any individual child in terms of the number of units *below* x. The age at which the curves for the individual tests coincide can be controlled by the height at which the various curves are plotted; but, no matter at what age they are made to coincide, the shape of the combined curve will be the same.

We do not attempt to determine the absolute height of any portion of the intelligence curve. We merely attempt to plot a portion of it—that lying between the ages 6.5 and 13.5—in terms of strictly comparable units. This portion is, as it were, suspended in mid-air at an unknown height. If we knew its height, we would need no x. It does not matter, however, what point we take as the point whose absolute height we call x, for we do know the number of units of height of any point in the obtained portion above or below any other point in it.

It should be clear, then, what weighting a test means in the present case. That test is weighted most whose curve tends most to give the combined curve a vertical direction. If, at certain points, it tends more strongly than at others to give verticality to the combined curve, it is weighted more at those points. On the other hand, that curve which tends most to make the combined curve horizontal is weighted least. With these considerations in mind, it will be clear that the easy opposites test has greater weight from one end to the other than the three letter cancellation test. The point-norm curve for easy opposites, being steeper than that for the cancellation test, tends more to give verticality to any combined curve into which it enters as a component. Further, it will be realized that in those regions where the curve for the cancellation test runs more nearly horizontal than at others, the test has less weight. Thus, between the ages of 9 and 10, where the curve for the cancellation test runs exactly horizontal, the test as a test for discriminating between the mental ages nine and ten has a weight of zero, though it retains a not incon-

siderable weight as a test for discriminating between these ages and those above or below them.

One more question which may arise in connection with the present scale is, why it is called an *absolute* scale? This can best be understood, perhaps, by recalling the facts which cause it to be generally conceded that the absolute increase in mental age from year to year varies, tending to decrease at the upper ages. This generally accepted view rests primarily on the fact that as we reach the higher ages, the percentage of children of one age able to pass the same tests as the average child of the next younger age tends to decrease. It is easy, for example, to find tests passed by only 50 per cent of three year olds, which can yet be passed by 80 per cent of four year olds, whereas it is difficult to find tests passed by 50 per cent of thirteen year olds which are passed by even 60 per cent of fourteen year olds. This variation in the percentage of children of one age who pass the same tests as the average child a year younger is taken to indicate variation in the absolute difference between any two adjacent mental ages.

In short, it is generally conceded that the difference in these percentages serves as an index to the absolute difference in mental ability between any two mental ages. Now the discriminative values which we have used are, as has been pointed out, values which correspond directly with the percentages mentioned; *they are the mathematical measures of that difference between the performances of two adjacent age groups which determines what percentage of the upper group will exceed the average of the lower;* consequently they are the mathematical measure of the absolute difference between the age groups, and a scale of which they constitute the units is an absolute scale.

The original score norms and their values in absolute intelligence units or points is given for all nine tests in Table III. The point values are all worked out in the way already described, illustrated in Table II.

TABLE III

VALUES OF ORIGINAL NORMS WHEN TRANSFORMED INTO ABSOLUTE INTELLIGENCE UNITS

Av.=untransformed norms. Points=value in absolute intelligence units.

Age		6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5
Easy Opposites..	Av.	0.0	1.1	2.6	4.4	5.2	6.5	7.4	8.6
	Points	x	0.9	2.0	2.9	3.2	3.9	4.3	4.9

Hard Opposites..	Av.	0.0	0.5	1.3	2.8	3.7	6.3	7.1	9.1
	Points	x	0.5	1.1	1.8	2.1	2.8	3.0	3.4
Substitution.....	Av.	10.1	20.5	30.2	40.3	44.2	50.6	52.3	59.1
	Points	x	1.2	2.1	3.0	3.3	3.8	4.0	4.6
Word Building.	Av.	0.4	2.3	4.3	6.3	7.4	9.8	10.3	12.0
	Points	x	1.1	1.9	2.5	2.8	3.5	3.6	4.1
Digits' Span....	Av.	3.0	4.8	4.9	5.3	5.4	5.7	5.9	6.0
	Points	x	1.2	1.9	2.4	2.6	3.0	3.2	3.3
Language Completion	Av.	1.7	5.0	7.5	10.3	11.4	14.2	15.0	15.1
	Points	x	1.3	2.1	2.9	3.2	3.9	4.6	4.1
Anagrams.	Av.	1.8	4.5	7.4	9.1	9.2	10.2	11.1	11.7
	Points	x	0.9	1.8	2.3	2.3	2.7	3.0	3.2
Cancellation...	Av.	12.1	16.0	17.2	20.6	20.6	23.0	23.7	25.6
	Points	x	0.8	1.1	1.9	1.9	2.4	2.5	2.8
Comprehension..	Av.	2.5	4.0	5.8	9.5	13.1	16.1	16.8	17.2
	Points	x	0.7	1.3	2.0	2.6	3.3	3.5	3.6
Total.....		x	8.6	15.3	21.7	24.0	29.3	31.2	34.0

(i. e., absolute intelligence scale.)

Point Values for any Possible Score: After the number of points corresponding to each age norm, as shown in Table III, had been worked out, the next step was to work out the number of points corresponding to any possible score. This was done by the same formula as that used in finding the number of points corresponding to the norms. The difference between a given possible score and the next lower norm was divided by the average of the standard deviations of the two norms between which the score in question occurred. The result, added to the number of points corresponding to the next lower norm, gives the point value of the score in question. Points were assigned in this way for every possible score lying between the norms for 6.5 yrs. and 13.5 yrs. inclusive.

As we had no norms beyond these age limits, the problem of how to deal with scores falling beyond the norms for these ages was a difficult one. It was finally decided to reckon the number of points below x corresponding to any score below the average of the 6.5 year group in terms of the standard deviation of that group; and, similarly, to reckon the number

of points above the 13.5 year point norm corresponding to any score above the 13.5 year average in terms of the standard deviation of the 13.5 year group. Deviations below the 6.5 year norm, then, were transformed into points by dividing them by the standard deviation of the 6.5 year group, while deviations above the 13.5 year norm were divided by the standard deviation for that age group and the resulting number of points was added to the number of points corresponding to the 13.5 year norm. Since x, or arbitrary zero, is placed at the 6.5 year average, the point values of scores below x are negative. They were worked out for each test down to the point value of a performance score of zero.

This procedure used in calculating the point values of scores lying above our highest norm, or below our lowest, differs from the procedure used elsewhere in the scale in that elsewhere the differences are divided by the average of two standard deviations, namely, the standard deviations of the norms both below and above the score in question. Since we have no norms, or averages, below 6.5 or above 13.5, it is necessary with scores respectively below and above these points to use as a divisor simply one standard deviation,—that of the 6.5 year group for scores below the 6.5 year norm and that of the 13.5 group for scores above the 13.5 year norm. Such a procedure involves the assumption that the 6.5 year standard deviation applies to all scores below the 6.5 year norm, and the 13.5 year standard deviation to all scores above the 13.5 year norm. There is an element of error in this assumption, as standard deviations tend to vary with age. It is possible therefore, that by using the 6.5 year standard deviation, we have penalized failure to reach the 6.5 year norm less severely than if the correct standard deviations of the lower age groups could be taken into consideration. For scores above the 13.5 year norm, however, it is not likely that there is any appreciable error involved, since the standard deviations tend not to change much after the 13th year.

We thus worked out a table of nine columns, giving the number of points to be assigned to any possible score in each of the nine tests. Four of these columns are given by way of illustration, in the accompanying table, No. IV. Scores lying outside of the horizontal lines lie outside of the obtained age norms, and were calculated as explained in the preceding paragraphs.

TABLE IV

POINT VALUES FOR ANY POSSIBLE SCORE

Hard Opposites		Language Completion		Comprehension		Anagrams	
Score	Points	Score	Points	Score	Points	Score	Points
0	x	0	—0.8	0	—0.8	0	—0.8
1	0.9	1	—0.3	1	—0.6	1	—0.4
2	1.5	2	0.1	2	—0.2	2	0.1
3	1.9	3	0.5	3	0.2	3	0.4
4	2.1	4	0.9	4	0.7	4	0.8
5	2.4	5	1.3	5	1.0	5	1.1
6	2.7	6	1.6	6	1.3	6	1.4
7	3.0	7	1.9	7	1.5	7	1.7
8	3.2	8	2.2	8	1.7	8	2.0
9	3.4	9	2.5	9	1.9	9	2.3
10	3.6	10	2.8	10	2.1	10	2.6
11	3.8	11	3.1	11	2.3	11	3.0
12	4.0	12	3.3	12	2.4	12	3.3
13	4.2	13	3.6	13	2.6	13	3.6
14	4.4	14	3.8	14	2.8	14	3.9
15	4.6	15	4.1	15	3.0	15	4.2
16	4.8	16	4.3	16	3.2	16	4.5
17	5.0	17	4.5	17	3.6	17	4.8
18	5.2	18	4.7	18	3.9	18	5.1
19	5.4	19	4.9	19	4.2	19	5.4
20	5.6	20	5.1	20	4.5	20	5.7
		21	5.3	21	4.8	21	6.0
		22	5.5	22	5.1		
		23	5.7	23	5.4		
		24	5.9				
		25	6.1				
		26	6.3				

Absolute Intelligence Scale: The quotients which we have termed discriminative values represent an absolute increase in ability from one age to the next, for their respective tests. If the nine tests used can be regarded as tests of intelligence, then a scale built up from an amalgamation of the results of all nine may properly be regarded as a scale of absolute increase in intelligence (beyond a starting point, x, arbitrarily taken as the norm for 6.5 yr. old children). The final absolute intelligence scale is thus secured simply by adding together the point scales of the individual tests. The normal absolute intelligence score for any age is simply the sum of the normal point scores for that age of all nine tests. This has already been given in Table III (bottom line).

The absolute intelligence scale represents the normal growth

in absolute intelligence from the age of 6.5 to 13.5 years. We have plotted this scale as an intelligence growth curve, in Figure 2. It will be observed that the curve brings out the fact that growth in intelligence is considerably more rapid in the earlier years than in the later ones. It looks as if, were it continued at the upper end, it might reach a level somewhere in the neighborhood of 16 or 17 years, and, on the other hand, if continued at the lower end, become much steeper at the first few years of life.

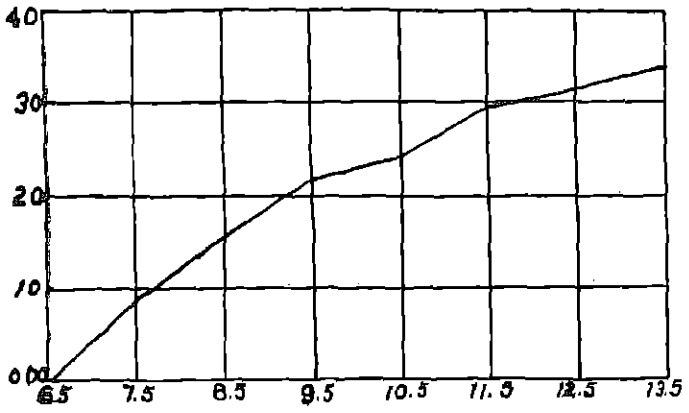


FIGURE 2.—Curve showing growth in intelligence, measured in absolute units, between the ages 6.5 and 13.5. Ages are represented by abscissæ, absolute intelligence units by ordinates.

As already stated, to determine the total point score for each age, the points corresponding to the norms of each age group for the nine tests were added together. Norms for months were then interpolated between these year norms. The final scale, showing the number of absolute intelligence units corresponding to each age, from 6.5 years to 13.5 years, inclusive, by months, is given in Table V.

TABLE V

NUMBER OF ABSOLUTE INTELLIGENCE UNITS CORRESPONDING TO EACH MONTH OF AGE FROM 6.5 YEARS TO 13.5 YEARS

	0 mo.	1 mo.	2 mo.	3 mo.	4 mo.	5 mo.	6 mo.	7 mo.	8 mo.	9 mo.	10 mo.	11 mo.
6 yrs.	....	....	....	....	....	....	....	x	.7	1.4	2.2	2.9	3.5
7 yrs.	4.3	5.0	5.8	6.5	7.2	7.9	8.6	9.2	9.7	10.3	10.8	11.4
8 yrs.	12.0	12.5	13.1	13.6	14.2	14.8	15.3	15.8	16.4	16.9	17.4	18.0
9 yrs.	18.5	19.0	19.5	20.1	20.6	21.1	21.7	21.9	22.1	22.3	22.5	22.7
10 yrs.	22.8	23.0	23.2	23.4	23.6	23.8	24.0	24.4	24.9	25.3	25.8	26.2
11 yrs.	26.6	27.1	27.5	28.0	28.4	28.8	29.3	29.5	29.6	29.8	29.9	30.1
12 yrs.	30.3	30.4	30.6	30.7	30.9	31.1	31.2	31.4	31.7	31.9	32.1	32.4
13 yrs.	32.6	32.8	33.0	33.3	33.5	33.7	34.0	....	....	....	....	....

By the aid of Table V, mental ages may be determined and intelligence quotients may then be reckoned in the usual manner. The procedure involved, then, in measuring the intelligence of any member of a group of children, by the present method, is as follows: First, the original scores in each of the nine tests is determined. By the use of a nine column table these original scores are changed to point scores. The sum of these nine point scores is the number of units of intelligence credited the child. This number of units can be immediately transformed into mental age by the use of Table V.

Results of Application of the Scale: Measurements of the absolute intelligence above the average for 6.5 years were secured for 490 pupils. These measurements were then converted, by the use of Table V, into terms of mental age and intelligence quotient. It is usually held that the first criterion of the reliability of any scale for the measurement of children's intelligence is that the median child of each age group should be found to have a mental age just equal to his chronological age. Now our norms are based on averages and not on medians, and it is fairly clear that the average mental age of our various chronological age groups runs higher than their median. For most of our groups, however, it is impossible to determine the exact average mental age, for in many cases our only information is that the mental age is over 13.5 or less than 6.5. Consequently, in place of a table comparing average chronological ages with average mental ages, such as we should prefer to give, we must be content with the usual comparison of median ages. By the aid of Table VI, it may be seen how closely the

TABLE VI

MEDIAN MENTAL AGE FOR EACH AGE GROUP

Chronological Age of Group	Median Chronological Age	Median Mental Age
7.5 yrs.	7 yrs. 6.5 mos.	7 yrs. 4.5 mos.
8.5 "	8 " 6.75 "	8 " 4.6 "
9.5 "	9 " 5.4 "	9 " 3.2 "
10.5 "	10 " 6.4 "	10 " 4.5 "
11.5 "	11 " 7.1 "	11 " 5.0 "
12.5 "	12 " 6.2 "	12 " 1.3 "
13.5 "	13 " 5.0 "	13 " 2.5 "

median mental ages parallel the median chronological ages. While the median mental ages run uniformly a month or so below the median chronological ages, they always fall well within the year covered by the ages of the corresponding

chronological groups. So far as this criterion is concerned, we regard our tests as satisfactorily "standardized," though actually we consider them as rather poorly standardized, because we have evidence from the age and school-grade distributions that the different age groups are not equally representative.

As further evidence of the general accuracy of the present scale, we may cite the correspondence between measurements made by it with those made by individual testing. For the purpose of making this comparison, 51 children were tested individually by Kuhlmann's latest revision of the Binet tests. The degree of correspondence is shown, in terms of intelligence quotients, in Table VII. The uncorrected coefficient of

TABLE VII

COMPARISON OF INTELLIGENCE QUOTIENTS OBTAINED BY GROUP TESTING WITH THOSE OBTAINED BY INDIVIDUAL TESTING WITH THE KUHLMANN REVISION.

Chronological Age	Intelligence Quotient		Chronological Age	Intelligence Quotient	
	Group	Individual		Group	Individual
7— 5	113	111	9— 3	114	114
7— 2	106	123	7— 8	121	111
7—10	83	85	7—11	108	116
7— 5	93	98	7— 7	115	117
7— 8	105	121	9— 0	117	124
13— 7	88	99	9— 7	75	75
13— 2	67	73	8— 7	100	102
12— 5	91	82	8— 6	91	102
12— 0	97	98	8— 7	107	117
10—11	79	80	8— 1	94	107
10— 3	79	80	8— 9	93	91
10—10	79	86	8— 8	100	103
10— 7	77	88	8— 3	96	100
10— 4	91	92	8— 0	108	111
10—5	91	93	8— 4	105	101
10—10	103	94	8—10	109	105
10— 6	93	91	8— 6	96	104
10— 7	97	114	8— 1	105	119
10— 8	100	100	8— 9	90	88
10— 1	93	98	8— 2	96	93
10— 0	88	82	8— 7	86	102
10— 2	83	87	7— 1	101	128
10—11	68	77	7— 9	95	103
9— 5	124	124	7—	106	111
9— 0	96	97	7— 7	120	134
9—11	88	92			

correlation between the two sets of quotients is 0.78 (Bravais-Pearson). In view of the greater accuracy of individual tests as compared with group tests, the magnitude of this correlation is rather surprising — particularly so when it is considered that our tests are probably not the best that could be selected and that our norms are not to be regarded as finally perfected. We incline to contribute a considerable portion of the apparent success of the present tests to the method we have used, and to believe that inadequacy of the tests and norms is compensated by the advantages inherent in the methodological principles we have followed.

A further interesting comparison between the results obtained in the present study and those of other investigators is afforded by Table VIII, which gives the distribution of cases

TABLE VIII

DISTRIBUTION OF THE MENTAL AGES OF EACH CHRONOLOGICAL AGE GROUP

Mental Age	Chronological Age 7:0 to 7:11	8:0 to 8:11	9:0 to 9:11	10:0 to 10:11	11:0 to 11:11	12:0 to 12:11	13:0 to 13:11	Total Cases
Less than 6:6	6		1					7
6:6— 6:11	16	1	..	..	..	..	..	17
7:0— 7:11	30	19	6	2	1	..	..	58
8:0— 8:11	17	34	21	11	1	1	2	87
9:0— 9:11	3	13	18	23	11	3	3	73
10:0—10:11	..	3	7	17	18	15	3	63
11:0—11:11	..	1	4	18	14	17	5	59
12:0—12:11	..	..	7	3	8	8	7	33
13:0—13: 5	..	..	1	1	9	5	9	25
More than 13:5	..	..	1	4	19	25	19	68
Total Cases	72	70	66	79	81	74	48	490

by mental age. At the ages of 7.5 and 8.5, the mode and median coincide, and the distribution appears fairly symmetrical. At all of the upper ages, however, with the exception of the small and unrepresentative 13.5 year group, the distribution of mental ages is skewed toward the lower end. Even the 7.5 and 8.5 year groups show a very slightly skewed distribution, when the exact mental ages by months are considered. This is seen, for example, in the fact that the average mental age of the 8.5 year group is 8 yrs. 5.5 mos., while the

median mental age is slightly less, namely, 8 yrs. 4.6 mos. The number of cases is too meagre to justify final conclusions. At present, we may simply note the fact that our data suggest that the distribution of mental ages varies with the chronological age of the group, being almost symmetrical at the ages of 7.5 and 8.5, but becoming rather definitely skewed at the higher ages.

This tendency toward skewed distributions offers an explanation of several minor matters in some of the preceding tables. Since in a skewed distribution it is natural for the median to fall below the average, and since our norms are based on averages, it is possible to understand why, in Table VI, the median mental ages are slightly lower than the median chronological ages. Skewed distributions also account, in part, for the fact, brought out by Table VII, that the present tests very frequently give intelligence quotients which are lower than those obtained by the Binet tests, while at the same time they show, particularly at the upper ages, as seen in Table VIII, an unusually large number of children with a mental age well above their chronological. Pintner, using group tests, likewise found a number of cases showing a mental age greatly above their mental age by the Binet tests, though the median mental age of the group with the group tests was a quarter of a year less than the median with the Binet tests.[10] In a skewed distribution, it is easily possible for a heaping up of cases at the lower mental ages to be accompanied by a considerable number at the more advanced mental ages.

[10] *The Mental Survey*, 1918, pp. 61-63.

"CROSS-OUT" TESTS

WITH SUGGESTIONS AS TO A GROUP SCALE OF THE EMOTIONS

STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

By S. L. Pressey and L. W. Pressey

The tests described in the present paper are all "cross-out" tests—that is, each one asks of the subject that by crossing out some one thing he eliminate a wrong, irrelevant, or extreme element in a situation.[1] The writers cannot avoid the feeling that the great stress at present being put upon rigid systematization in method and objectivity in scoring has resulted in not a few tests of form so artificial and unnatural as to seriously interfere with their usefulness. The notion of the elimination of an extra or wrong element was hit upon by the writers as a way in which systematic method and scoring might be obtained, without any such sacrifice of rationale or naturalness in the problem, as presented to the subject. To cross out a mistake is a very natural thing to do. It is a child's own method, as any teacher who has had to train children to erase *instead* of crossing out, can testify. From the subject's point of view the task presented in the tests is thus not at all unreasonable. Such a framing of the problem is, in the second place, satisfactory from the point of view of the examiner, since the directions can be very brief and direct. Finally, so far as scoring is concerned, the plan is admirable; either the wrong word or part is crossed out, or it is not. There is no writing, even of numbers. Indication of the answer by underlining is not wholly ideal (while being at the same time an unnatural and artificial way of answering a question): not infrequently a line is put half way between two words—the writers have found subjects who put the line

[1] The general principle involved in the tests, and certain of the tests described in this paper, have already been mentioned briefly in a previous article, Sex Differences Shown by 2544 School Children on a Group Scale of Intelligence, with Special Reference to Variability, Pressey, L. W., *Journal of Applied Psychology*, Vol. II, 1918, pp. 323-340.

over the right word. But if the line is drawn *through* the word there is a minimum of possibility of misinterpretation.

The writers have now tried this principle of the "extra" or "wrong" element in a variety of tests, with a variety of subjects, and have found it to work excellently. The present paper will describe briefly (1) a brief group scale of intelligence, to be used in school surveys, and applicable from the fourth through the eighth grade, and (2) a scale supplementary to this, for use from the kindergarten through the third grade; (3) a group scale for measuring emotional interests will be outlined, and suggestions made as to a test of learning, a test of ingenuity with geometrical forms, and a silent reading test. There is here surely variety enough; but all these tests make use of this same "cross-out" principle. And one might almost say that a single brief direction would serve for all: "Cross out what is wrong, or does not belong with the others."

I. Survey Scale Unit E

The blank is partly reproduced below, only the first 5 of the 25 items in each test being given here. Briefly the directions are as follows: in the first test the children are told that each list of words is a sentence with the words mixed up. But there is in each list one word that does not belong in the sentence. This word they are to cross out. In the second test the children are told simply to cross out in each list the thing that does not belong with the other four things. For the third test it is explained that the numbers in each list are arranged in accordance with some rule, but there is one number in each list that breaks the rule; and this recalcitrant number they are to cross out. Finally, in the last test the children are told very simply to "Cross out in each list the thing that is worst." Five minutes are allowed for the first and third tests, four minutes for the second and fourth.

The results below are from 850 school children (grades four to twelve inclusive) in an Indiana city of about 12,000 population. By total score on the four tests the results are as follows:

Grades	4	5	6	7	8	9	10	11	12
Number	178	147	111	89	68	102	56	42	57
Medians	40.5	48.1	56.1	60.0	69.5	71.1	72.0	74.0	75.5
25%	32.7	43.1	48.1	55.1	64.0	66.4	67.5	64.2	72.6
75%	47.2	54.2	60.2	68.2	75.0	74.6	77.2	81.0	80.0

Results by the individual tests were for each grade as given in Table I. It will be seen that there is a fairly good progress

Schedule E.

MENTAL SURVEY SCALES

"CROSS-OUT" TESTS

Name ..... Sex .....
Age ..... Birthday ..... Grade .....
Teacher ..... School .....
Place ..... State ..... Date .....

TEST I. VERBAL INGENUITY

Examples: (a) see a I man on.
(b) knife chair the sharp is.
(c) cat dog the house after ran the.

(1) the cat at see.
(2) boy was sky the sick.
(3) bread sweep will the kitchen I.
(4) are going yesterday to-morrow we.
(5) me mine give my straw hat, etc.

TEST II. LOGICAL JUDGMENT

Examples: (a) dog, cow, horse, oak, cat.
(b) book, newspaper, magazine poster, photograph.
(c) chair, desk, table, wagon, book-case.

(1) coat, shoes, hat, gloves, cow.
(2) oats, wheat, barley, cotton, rye.
(3) satin, silk, linoleum, calico, serge.
(4) fly, ant, bee, grasshopper, mouse.
(5) bread, meat vegetables, hay, fish, etc.

TEST III. ARITHMETICAL INGENUITY

Examples: (a) 2 4 6 8 9 10 12
(b) 7 6 5 6 2 7
(c) 1 4 7 9 10 13

(1) 1 2 3 9 4 5
(2) 2 4 6 7 8
(3) 19 18 17 16 13 15
(4) 5 7 10 15 20 25
(5) 11 10 8 6 4 2 etc.

TEST IV. MORAL JUDGMENT

Examples: (a) Gambling, lying, drunkenness, murder, smoking.
(b) begging, gambling, trading, cheating, borrowing.
(c) dullness, foolishness, laziness, weakness, poverty.

(1) fighting, killing, hating, quarreling, hurting.
(2) borrowing, gambling, overcharging, stealing, begging.
(3) love, hate, fondness, dislike, liking.
(4) dancing, drunkenness, flirting, over-eating, smoking.
(5) holiness, reverence, piety, obedience, wickedness, etc.

from grade to grade through the eighth (as far as the tests are expected to apply), with a level after that.

As regards the structure of these tests the writers wish to call attention to the following points: (1) the possibility of chance successes is at a minimum; in the second and fourth tests the chance is one in five, in the first and third, still less; (2) there is possible great condensation in the printing of the tests; an examination of one hundred clear-cut problems is presented on two sides of a single 12 in. by 9 in. sheet of paper. Such condensation is saving of space, trouble in scoring and handling, and expense of paper and printing; (3) the method by which the subject indicates his answer is the extreme of simplicity; he simply crosses out a word. The manual labor of taking the examination is thus at a minimum; and the rate at which the subject progresses is determined solely by his ability to solve the problems. Finally (4) scoring is extremely easy, rapid, and wholly objective. There is in each line only one correct answer. The subject has indicated his response to each problem not by writing, nor even by so simple a thing as a number, but by drawing a line through a certain word or number, one only in each line. It may be said that the examination takes less than 25 minutes to give, and the blanks can be scored at the rate of 70 an hour by student help.

SURVEY SCALE UNIT E

RESULTS BY TEST

Grade	Test I. 25%	M.	75%	Test II. 25%	M.	75%	Test III. 25%	M.	75%	Test IV. 25%	M.	75%	Grade
4	8	10.7	14	10	11.9	14	7	9.7	13	7	9.4	11	4
5	11	13.6	16	11	13.0	15	9	12.1	14	10	11.6	13	5
6	13	16.6	18	12	14.0	16	11	13.1	16	11	13.0	15	6
7	15	17.7	20	14	16.1	18	12	14.9	17	13	15.4	17	7
8	17	19.5	21	17	18.7	20	14	16.5	18	16	17.3	19	8
9	18	20.1	22	17	18.5	21	14	16.2	18	15	17.3	18	9
10	18	21.5	23	18	19.6	21	15	16.7	18	16	18.2	20	10
11	16	20.0	23	17	19.6	21	14	16.8	18	17	19.4	21	11
12	19	21.1	23	18	19.4	21	15	16.9	18	18	20.5	22	12

But the writers wish to emphasize particularly certain features of the examination as it appears from the point of view of the subject. There is to each test a rationale and a plausibility that immediately arouse both interest in and understanding of the test, with only brief directions. The subject's response—the crossing out of an extra, incongruous, or undesirable element—is so natural as to be considered a matter of course. Further, the directions are in essence the same for all four tests, and the subject's response is the same. It may be said that for 850 subjects, from the fourth grade on, there

were obtained no zero scores on the total examination. There were only 8 zero scores on separate tests (5 of these on the third) and only 20 scores below 3 on any test (9 of these on the third).

The writers have been quite as much interested in the individual tests as in any ratings that might be obtained from the total examination. The first test might be expected to give a measure of language ability. As such it should have definite superiority over the Trabue Completion Scales in its absolute objectivity of scoring and its elimination of writing. It should in addition sample more complex mental processes, since the mental reconstruction of a sentence would seem a more elaborate effort than the mere completion of it. It should also be pointed out in passing that this form of "dissected sentences" is greatly superior to the form used in the army scale Alpha in its condensation of matter, its elimination of the possibility of chance success, and its elimination also of elements of information from the problem.

The second might be called a test of information with understanding.[2] It should be noticed that the test presents in each line a problem in both differences *and* similarities. The subject must make sure that four things in each list are alike, and the fifth different from *all* the other four. The Binet separates the problems and asks either "how are these two things different" or "how are these two things alike." Group test phrasings of this problem have been of two forms; in one the subject is presented with pairs of similarities or opposites, and is asked to underline either the "similar" or "opposite" as one or the other applies; in the other he is given a word, followed by a number of other words from which the opposite or similar is to be selected by underlining. The superiority of the present phrasing in compactness, naturalness and simplicity of direction would seem unmistakable.

The writers have hoped that the "arithmetical ingenuity" test might give a measure of potential arithmetical ability, and so give results in interesting contrast to results with tests of the fundamental processes in arithmetic, in classes where the teaching was poor but the abilities of the children good. The

[2] A test of this form for technical knowledge in a particular field would seem possible. A skilled mechanic might be given lists of various tools or processes, with in each list one process or tool not properly belonging there; four lathe parts and a milling machine part, for instance.

test in this present form, however, is not wholly satisfactory.[3] The product moments correlation for this test with results from a group scale of determined reliability previously described in this JOURNAL[4] shows for 205 6th, 7th and 8th grade children a coefficient of only .34 as compared with .61 on the first test, .65 for the second, and .91 for the last.

This high correlation of the last test with intelligence ratings otherwise obtained is extremely interesting. The test undoubtedly involves vocabulary, and general capacity for abstract thinking, as well as moral sentiment and judgment. In formulating the test the writers had in mind, of course, these last abilities especially, and they hope to obtain results on this test from the state reform school, to see whether any relation appears between ethical discrimination, as this test measures it, and delinquency. But altogether aside from the interest of its special content, the test has, apparently, a surprisingly high value as a measure of "general ability."

Mental Survey Unit E is thus a scale of four tests, each test of independent interest, the combined result giving also something of a rating of "general ability." The writers have thought of the scale as of value chiefly for the purpose of comparing classes, schools, communities, or other groups. As they have shown elsewhere,[5] such comparisons may be of distinct value in the study of educational problems. They hope to present shortly data bearing on the sociological value of such group measurements. And they feel that for such investigation "research units" of the general nature and form of this small four-page examination, taking approximately 20 minutes to give and scored with extreme readiness, are especially satisfactory.

II. SURVEY SCALE UNIT F

Samples of the four tests of the examination are shown below:

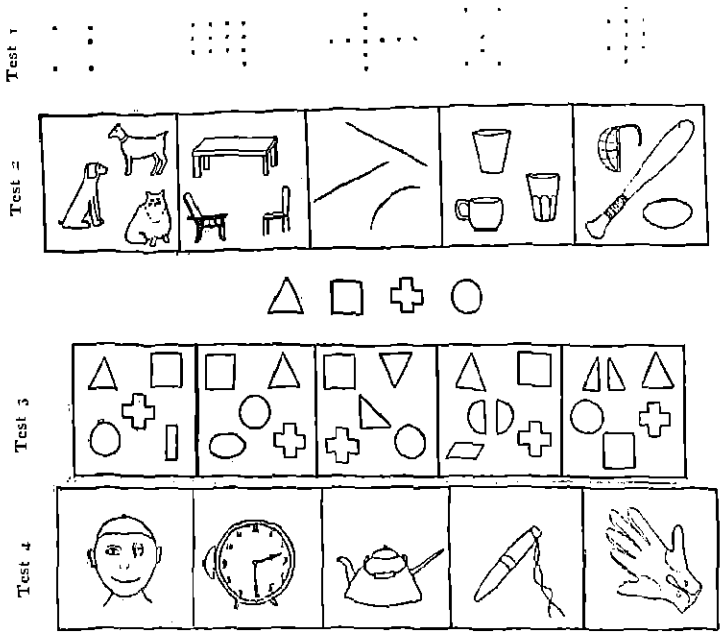
---

[3] The test is very poorly graded, and involves too great a variety of problems. It is now being revised so that the "numbers count up or down, multiply or divide in some regular way;" such irregular progressions as are shown in the thirteenth item are dropped altogether.

[4] A Group Point Scale for Measuring General Intelligence. *Journal of Applied Psychology*, Vol. II, 1918, pp. 250-269.

[5] Pressey, S. L. A Comparison of Two School Systems by Means of a Group Scale of Intelligence, *Educational Administration and Supervision*, Vol. V, 1919, pp. 53-62.

PLATE I



In the first test the children are asked to cross out the extra dot; in the second to cross out the thing that doesn't belong with the other two; in the third to cross out the block left over, in each square, after the other blocks have been fitted into the four forms (square, triangle, cross and circle) at the top of the page; in the fourth test to cross out, in each picture, the part that is wrong. Each test contains 25 items (with 5 items additional for examples). The children are allowed three minutes for each test. Results from 470 children in the 1st, 2nd, and 3rd grades of the same system surveyed by means of scale E show the following results:

Grades	1	2	3
Number	165	158	147
Medians	34.7	48.3	57.0

The medians for the separate tests at each grade were as below:

Grade	Number	Tests I	II	III	IV
1	165	5.5	10.9	6.4	13.3
2	158	10.4	12.4	10.6	17.0
3	147	13.3	13.6	14.0	18.4

It will be seen that the tests are not as yet properly graded, but the general principle of each one of the four tests has proved satisfactory. All four tests can be "gotten over" even to kindergarten children, and indications to date would lead one to expect a high correlation with intelligence as soon as the problems are properly graded.

The problem faced, in building these tests, was not an easy one. It was desired (1) that the tests should not require literacy, not even so much as knowledge of the alphabet or the cardinal numbers; otherwise more or less incidental influences such as early home teaching and training might play a disproportionate part in the results. The difficulty was to present *without* the use of any such symbols, problems of such a complexity as would give a measure of general intellectual ability. It was also desirable (2) that the children should indicate their answers in some extremely simple and unmistakable way. Young children show too poor eye-hand coordination to ask them to complete a picture, follow through a maze, or go through similar complicated processes, in the group, without greatly lengthening the examination, wearying the children, and introducing difficulties of scoring which would make the examination almost impracticable. Finally (3) it became evident very early in the experimentation that very young children are as yet so little disciplined in habits of attention that directions of any length, or problems of any artificiality or lacking in intrinsic interest, were not to be thought of. The directions must be almost colloquial in their simplicity and directness; the test must, in fact, be self-explanatory. And the examination must have more the appearance of a game, than of a task to which there must be conscious application.

The "cross-out" idea seemed admirably suited to the problem. The dot pattern test was tried with some hesitation. The results show the items to be poorly graded, but the problem of the test was better understood by the child than was expected (and this in a system without any kindergarten training which might afford some special preparation for such a test). So far as could be judged from the data now in hand, there is a fair correlation with general ability.

Test II involves the same problem as Test II of the previous scale (in picture form) and needs no further comment. Test III is an attempt to put a form board test on paper, and in such a way as could be used in the group. It is the result of a long series of experiments and has in it possibilities of further development which will be returned to shortly. The last test, which is obviously the best of the four technically, permits the presentation of an indefinite variety of problems, and of a range of hardness from 98% success in the kindergarten to problems difficult enough to puzzle an intelligent adult.[6]

But the interest in the Primer scale has been not so much in the separate tests as in the total rating. The scale is an attempt to make possible a survey of the *entire* pupil material of a school or school system. As a matter of actual fact the data thus far presented is the result of such a survey of *all* the school children, *from the first grade through High School,* in a city of about 12,000 population.[7]

The writers have already[8] presented data to show that schools may differ markedly as regards the grouping which they make of the children, as regards ability, in grades and sections. A survey which shall include the *entire* school system is thus necessary if an adequate picture is to be obtained of the school's disposal of its pupils. And a measure of the abilities of the children before such selective and formative influences have had much opportunity to operate is necessary, if a fair estimate of the community is to be obtained. This primer scale should then, if found satisfactory, give results of the greatest sociological as well as educational significance In fact, for all such group comparisons it should give data of unique interest.[9]

---

[6] The test has interesting possibilities, as a trade test for example. A picture showing a piece of machinery with some part out of order—this part to be crossed out—would be excellent for such a purpose.

[7] It might be added that the total 1,321 children were tested in a day and a half, the primer scale being given exclusively by one of the writers and Scale E throughout by the other, and that the entire 1,321 blanks were scored by student help at a total cost of $8.00.

[8] See note 5 above.

[9] It may be said, shortly, that children entering the schools located in the best neighborhood of the city surveyed averaged over a year in "mental age" above children (of the same chronological age) entering another school in a poorer section of the city, and that the median rating of grade 1 in the best school was approximately the same as the rating of grade 2 in the poorest.

III. Suggestions as to Other Tests, Especially of the Emotions

The following tests have been only superficially experimented with, with small groups of school children or college students. There is thus no systematic data to present. But the tests are of some intrinsic interest. The writers are hardly likely to find time for further work with them for some time to come; and since they are based on the same "cross-out" principle, they are described briefly here.

No one can have had any intimate experience with the feebleminded or the insane without feeling very strongly at times the almost pitiful inadequacy of a test of "general intelligence" as a single means of measuring social and economic inadequacy. One very important element in such cases is lack of emotional control. The writers wish to suggest the following four tests as making an interesting research unit for experimentation in the field.

1. The first test (to be placed on the first page of a four-page folder) consists of a story, beginning in very monotonous and colorless fashion, but in the middle of the page developing into a hyper-adventurous train robbery episode of the dime novel type. Sown throughout the test are irrelevant words, 20 in the first section and 20 in the second. The directions are to read the passage as quickly as possible, crossing out all the irrelevant words. The score consists of the difference between the number of words correctly crossed out in the exciting, as compared with the uninteresting section.

2. The second is a group adaptation of the familiar association test. There are 20 brief paragraphs of pied type on the blank. The subjects are told that 20 words will be read to them; that as soon as the first word is named they are to cross out all the letters in the first paragraph which occur in this word, and so on through. The first ten words are not emotionally toned; the last ten, such as are likely to touch a "complex" or stir up some emotion. And the score is the difference between the letters correctly crossed out in the two groups.

3. The third test consists of twenty-five lists of words, five words in each list. Some of these words are, for the average person, highly emotionally toned, some slightly and some not at all affectively colored. The subjects are told to go through the lists and cross out *all* the words that are unpleasant. After this is done, they are told to go through the lists again and, in each list, draw a line around the word that is *most* un-

pleasant. The slight experimentation with this test would suggest that the number of crossed words gave something of a measure of a person's tendency to emotionalize — or of affective spread. It is hoped that the number of deviations, in the selection of the most unpleasant word in each line, from the most common selection, may give a measure of neurotic "transvaluation."

4. The last test consists of a list of about 150 words. Of these half occur in the previous three tests, half do not. Of the words occurring in the previous tests, half are words with strong emotional coloring and half are words with no coloring. And of the words in the list, not appearing in the previous tests, half are emotionally toned. The subjects are told to go through the list and cross out the words that they have seen before, in the other tests. The record is scored as to (a) the proportion of emotional to unemotional words remembered and (b) the proportion of emotional to unemotional words remembered wrongly.

It has been the hope of the writers that they might find time to work out this four test unit with some care and obtain for purposes of comparison data from school children, a group of institutional feebleminded, and representative groups from the boys' and girls' state reform schools. No such results have been obtained as yet. But the writers are convinced, after a somewhat unusually varied and intensive experience in work with both the insane and psychopathic and the feebleminded, that in such extreme cases, at least, emotional instability is a gross enough thing to be measurable by some such means. And they feel very strongly that a study of the problem by *group* comparison is particularly needed. Individual differences thus drop out and it is possible to get at type reactions.[10]

When there is any considerable foreign element in a community tests intimately involving language can hardly be adequate. Other material is then necessary (such material may very well prove indispensable in all cases). That is, a group performance scale supplementary to scale unit E is needed. As two tests for such a scale the writers have experimented with (1) a puzzle test similar to the form board test above, showing at the top of the page a large square, and below in lines, groups of blocks, those in each group just sufficient to

[10] In fact they have an idea that such tests might get at a person's emotional peculiarities better when that person felt himself sheltered by being submerged in a group than would be the case in an individual examination.

fill the square but with one block left over. The extra block is, of course, crossed out. The writers have already obtained results from about 200 children and find the test highly satisfactory. As a second test (2) the writers have already in trial form a variation of the digit-symbol test. It consists of the usual blank, but with the substitutions already made, and in each line one substitution made wrongly. The children are told that in each line there is one wrong substitution, and they are to cross it out. The saving of time and labor in taking the test, and of trouble in scoring, over the standard form is obvious.

Finally the writers wish to suggest a new form of silent reading test employing the cross-out principle. Silent reading tests seem to fall into one of two pitfalls, either the pupils' performance cannot be objectively and satisfactorily scored or else his performance is more or less artificial, and with only a somewhat indirect bearing on the passage read. The writers have experimented tentatively with a silent reading test having such items as: "In winter the wind is very cold; it blows against our house, and often finds its way in through the cracks and keeps the house warm." The directions are to cross out the one word in each sentence that spoils the meaning. The problem as thus phrased is easily presented to the pupils, the scoring is wholly objective, and labor in both giving and scoring of the test is slight. The problem presented seems, to the writers at least, much less artificial than is usually employed in such tests and much more truly a test of comprehension of the meaning of the passage read.

IV. Group Tests as a General Method of Research in Psychology

A word remains to be said in defense of the presentation, in a single paper, of such heterogeneous material, much of it still in only experimental form. The writers have brought these various bits of experimentation together in just this way because they wish to emphasize the almost bewildering variety, in material, in method, and in problem, possible with the "group test." The tentative suggestions regarding four tests of the emotions will very likely be taken by the reader as amounting to little more than an opportunity for parlor amusement. For the writers these few paragraphs are perhaps the most serious part of the paper. They represent an attempt to formulate—in group test form—certain materials for dealing with one of the most baffling problems of psychology. The

"tests" are not intended as a means of diagnosis, but rather as a method of investigation. The writers believe that the majority of psychological problems can be attacked with profit by the "group test" method. They feel that it would be very unfortunate if the methods brought to general notice by the army psychological work, of investigation in the group, of systematic condensation of matter and record, and studied economy of time and labor in giving and scoring, should be limited in their application to the study of "general intelligence." Particularly are studies of groups *as such* likely to prove fruitful; we should thus be able to determine types, and to relate psychological study to large educational, sociological, and economic facts. The "group test" has appeared in the army work as a new method for dealing with certain problems of general intelligence; in reality it is a new method of investigation, applicable to the entire field of psychology.

Summary

The paper may be briefly summarized as follows:

1. The present emphasis upon rigid standardization in method and objectivity in scoring is tending toward tests of form so artificial and unnatural as seriously to interfere with their usefulness. As one way of obtaining greater naturalness—without sacrificing method—the writers suggest the formulation of test problems in such a way as to involve the elimination, by crossing out, of a wrong or unnecessary element.

2. First forms are presented of two brief group scales making use of this device; one for grades one to four inclusive, the other for grades four to eight,—the two together permitting the survey, for purposes of group comparison, of entire schools and school systems, from the kindergarten up.

3. Certain tests of special abilities, and a brief scale designed to measure emotional control, are suggested.

4. A plea is made for the recognition of the "group test" as a general method of investigation in psychology.

A COMPARATIVE STUDY OF THE VARIABILITY OF BOYS AND GIRLS

By George W. Frasier, Washington State Normal School, Cheney.

During the past decade much has been written and said about the variability of the sexes. One school of psychology maintains that there are no median sex differences of mental capacities, but the ability of the male is scattered over a greater range, that is, he is more variable. They thus account for the fact that more men than women have become famous, also that more men are in the low levels of society. The other school of psychology believes that there are no sex differences either in variability or central tendency. Furthermore they maintain that it is the present constitution of society that puts men into the extremes. The question of variability as applied to children is of vast practical importance to those who deal with boys and girls. Shall we deal with children on the assumption that boys *should* be found in the extremes and the girls in the middle ground?

The literature of the subject can be separated into two classes (1) studies dealing with men and women and making no mention of boys and girls and (2) those dealing with children. It is evident that many students of this subject draw their data from the behavior of adults, make their conclusions on such data and then apply their findings in dealing with children. This attitude is not justified by facts. The following study deals with school children, and for this reason the quotations taken from other authors will be confined to those dealing with boys and girls.

Keyes made a special study of all the children in one school district, having six or more annual records available. He found 1,902 children distributed as is shown in Table I.

TABLE 1

	Boys	Girls
Repeaters	373	310
Gainers	314	299
Normal	302	304
Total	989	913

He concludes from the above table [1] "The boys in this group (repeaters) are more numerous than the girls, as would be expected from the known greater variability of the male," also "The gainers or accelerates were 613 in number; 314 of them boys, 299 girls. The greater variability of boys is here again shown."

A casual glance at the table seems to vindicate the conclusion made by Keyes. However, it is to be noted that the totals show 76 fewer girls than boys. Now, Keyes has assumed that had these 76 girls been present all of them would have been in the "Normal" group. To overcome this inequality and show the true comparison between the boys and girls it is necessary to reduce each group to the percentage of the total. The percentage grouping is shown in table 2.

TABLE 2

	Boys	Girls
Repeaters	37.7%	33.9%
Gainers	31.6%	32.7%
Normal	30.5%	33.3%

The above table indicates that when these data are reduced to a comparable basis the greater variability of the boys vanishes.

Thorndike believes that boys are more variable than girls, for he writes [2] "Athough the male and female types are closely alike in intellectual capacities, there is an important difference in the deviations from the type in the two cases, namely, that the males deviate more. The highest males in any quality are more gifted than any of the women, and the lowest males inferior to all women. Thus, though girls in general rank as high or higher than boys in high school and college, they less often lead the class; thus there are far more eminent intellects among men than among women and twice as many idiots." In a foot-note he offers the following explanation: "This difference is not absolutely proven to exist but it appears fairly certain from the results of several investigations."

Rogers and McIntire in measuring the intelligence of children with the Binet-Simon scale found that [3] "The principal outcome is the familiar law that there is a wider scattering of ability among boys than among girls; consequently there are more gifted as well as more backward boys than girls," and

[1] C. H. Keyes. *Progress through the Grades of City Schools*, p. 15, 17 and 32.
[2] E. L. Thorndike. *Principles of Teaching*, p. 96.
[3] *British Journal of Psychology*, 1914, vol. 7, p. 265.

"so far as we go, our experiments support the general law, which several observers have already noted, that there is a greater variability of capacities in boys than in girls, but no general superiority."

The most complete study made in America dealing with the case in point was made by Terman. Terman studied 1,000 unselected school children from the standpoint of intelligence and came to the conclusion that [4] "The supposed wider variation of boys is not found. Girls do not group themselves about the median more closely than do boys. The range of I Q including the middle fifty per cent is approximately the same for the two sexes."

Leta Hollingworth made a very complete resumé of studies concerned with the variability of the sexes and pointed out that up to 1914 these had not established the greater variability claimed for the males. She says "—and there is at the present time no conclusive empirical evidence to show that in cases where the coefficient of variation is greater for one sex than for the other, this greater variability consists of greater range." [5]

It appears from the quotations given that the subject is far from settled, and for this reason the following data are presented.

It is claimed by the school of variability that whenever males and females, or boys and girls, undertake a task that the best progress and the slowest progress will be made by the boys, and the girls will compose the great bulk of median ability. Then, one would expect this law to hold true in school accomplishment.

It appears that should we select all of the boys and girls of a certain age in any school system or systems we could, by finding their grade location, determine the relative variability of boys and girls of that age.

Thirteen was selected as a desirable age because (1) children of this age are held in school by the attendance laws in most states and (2) at 13 a child has been in school long enough to demonstrate his ability.

The grade location of all thirteen-year-old boys and girls was found for twenty cities:[6] Boston, Mass.; Columbus, Ohio; Wichita, Kansas; Grand Rapids, Mich.; Kansas City, Mo.; Richmond, Va.; Winston-Salem, N. C.; Baltimore, Md.;

[4] L. M. Terman. *The Measurement of Intelligence*, p. 70.

[5] *American Journal of Sociology*, 1914., vol. 19, p. 510-530.

[6] The data concerning these children were taken from the city school reports.

Cleveland, Ohio; Trenton, N. J.; Philadelphia, Pa.; Danbury, Conn.; Fall River, Mass.; Winchester, Mass.; Beverly, Mass.; York, Pa.; Williamsport, Pa.; Aurora, Ill.; St. Louis, Mo.; Waterloo, Iowa. Table 3 gives the total number of each sex in each grade.

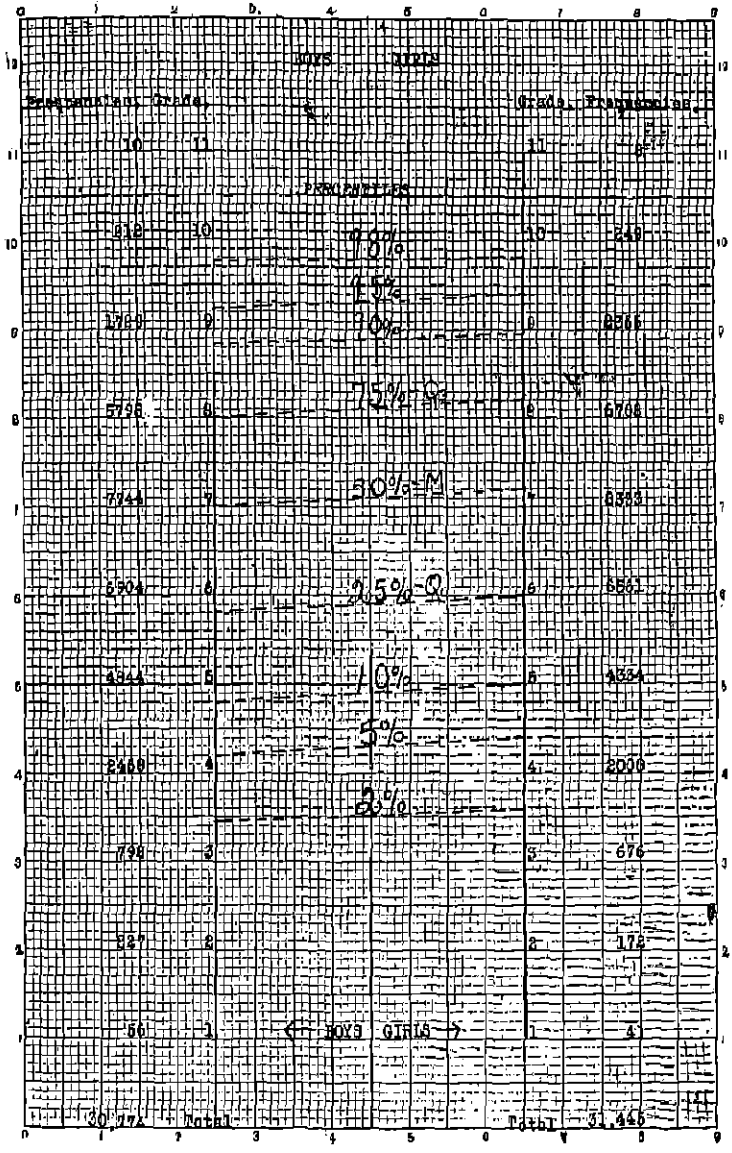
TABLE 3

Grade	Boys	Girls
I	56	41
II	227	172
III	792	676
IV	2,458	2,008
V	4,844	4,334
VI	6,904	6,561
VII	7,744	8,333
VIII	5,798	6,708
IX	1,729	2,355
X	212	249
XI	10	8
Total	30,774	31,445
Median	7.01	7.24
Q^1	5.81	6.10
Q^3	8.01	8.22
Q	2.2	2.12
C. V.	.16	.14

The coefficient of variability for the boys is slightly greater than that for the girls but the difference is so very small that no greater variability can be said to exist. In order to show all possible comparisons between the boys and girls chart A is given showing nine different percentiles, for each sex and their relative locations.

It is evident from the chart presented that girls make better school progress than boys. From the standpoint of variability our data point clearly to the conclusion that no sex differences are shown. The statistical procedure is carried out from the second to the ninety-eighth percentile so that the question of greater range is disposed of.

It is safe to conclude from the study of 62,219 thirteen-year-old boys and girls in 20 cities that the greater variability claimed for the boys is not present. It appears that this greater variability has never been established with any large number of cases or with any degree of certainty and if it is to be used as a working basis for those dealing with boys and girls it will be necessary for its supporters to present some conclusive empirical evidence. Until this is done it seems safest to disregard such a theory and deal with boys and girls on the supposition that they are alike so far as mental variability is concerned.


BOYS
GIRLS
PERCENTILES
1%
4%
10%
75%-M
50%-M
25%-M
10%
5%
2%
912
2769
5796
7944
6904
4344
2469
799
327
56
240
6868
6706
8333
6861
4334
2000
676
172
41
← BOYS GIRLS →
Total
30,774
Total
31,445

MENTAL ATTITUDES IN DEBATE

By Walter B. Swift, Boston

In speaking of mental attitudes in debate I am going to limit the subject to the moments of actual debate, to the few minutes when the speaker is on his feet before the audience. I shall have nothing to say about the time before or after the speech. What I have to say concerning development and change of mental attitudes will come later.

By "mental attitude" I mean "state of mind." Another name for mental attitude is "feeling," and another is "disposition." However, this feeling or disposition is not the normal feeling or disposition of the debater, the feeling that he holds generally, but is one which he holds during the moments when he is presenting his debate. It dominates him and dominates the situation at the time, but, unlike usual states of mind, it is not caused by the situation. On the contrary, it causes the situation, of itself. Again, the words which the debater is using do not cause the situation, but it dominates those words and gives them their effect. For example, suppose some one arises and says a few words. What he says may be clear, grammatical, and thoughtful. Yet the words themselves will not tell the whole story. There will be behind them a feeling, a state of mind — what I prefer to call a mental attitude — that will show through the words and impress itself upon the listeners. The words will mean more than they actually express, and they will do so because of the feeling that creeps into them and through the voice of the speaker.

Now as a rule, in the club of young debaters, most attention is given to the English and to an oratorical utterance. I have no fault to find with either sort of thing — indeed, I realize their value and I favor and encourage them. But I wish to emphasize the fact that this side of debating, which is a wholly external one, is not the only side to consider, and that possibly the other, the internal, mental one is after all the more important of the two.

Keeping this difference in mind, then, let me illustrate a little.

The usual way of reciting Yankee Doodle is somewhat like this: (Simple)

Yankee Doodle came to town
Riding on a pony.
He stuck a feather in his hat
And called him macaroni.

I can now so change my state of mind that an entirely different presentation is called for: (Sad)

Yankee Doodle came to town
Riding on a pony.
He stuck a feather in his hat
And called him macaroni.

Here let me call attention to one important thing: that each time I recited the stanza *the words were exactly the same.* Whatever change, therefore, was noticed in the reading was due to something else, to the change in my mental attitude as I recited the lines.

Here one may object that these two ways of reading are the only ones possible to this particular bit of verse. Let me answer by a third reading: (Revengeful)

Yankee Doodle came to town
Riding on a pony.
He stuck a feather in his hat
And called him macaroni.

These three ways of reciting the same words are sufficient to show that the change in the mental background may be a very varied thing, and that as the mental attitude changes, the rendering of the spoken word changes too.

Now what happened in my mind to cause this difference in my three readings? It was not a mere trick of the voice, but a real change that occurred much farther back than in the vocal mechanism, a change in the working of the mind that governs the vocal cords.

When I read Yankee Doodle the first time I did so in the usual way, giving the words the natural interpretation. But the second time I caused myself to think that Yankee Doodle's mother had just died; that he had come to town to attend her funeral; that he and his mother had been intensely dear to each other; and that her death was in consequence a great loss to him. This gave the recital an atmosphere of intense sadness.

At the third reading I pictured to myself the disastrous dealings of an enemy and the return of Yankee Doodle to the town for revenge. He had come to punish underhanded doings. Thus the reading reflected the change in mental attitude.

Just what I mean by mental attitude should now be clear. It is not personality, it is not quality of character, it is not a mental trait; it is rather an interpretation of the situation, it is the color or imaginative turn that one gives to his mind under the stimulus of a situation; it is the way that one deliberately makes himself feel in order that he may give force and an arousing power to his words.

Some may think that this is a little thing, that it is limited to little feelings or little mental attitudes which anybody can assume at will. I am going, therefore, to give two or three brief examples, in order to show the great variety and the possible power of these mental attitudes.

The first example will be the famous soliloquy in Hamlet, beginning with the phrase "To be, or not to be." Hamlet is at this point in despair and has come almost to suicide. At the verge he pauses a moment to think. So he is considering killing himself when he says "To be, or not to be." This phrase, when spoken in the first context, will sound quite different to any one whose senses are acute from that spoken in the second context. One easily realizes that each connotes quite a different thing. The feelings of the listener were different in consequence, and so were those of the speaker.

Let us take another example from Shakespere. Shylock says "I stand here on my bond." The situation was that he had just lost his ducats, that his daughter had eloped with a Christian, and that his ships had gone to the bottom. It must also be remembered that a Jew always desires full return on an investment and that he never makes one except upon that expectation. With these things in mind one will understand Shylock's feeling at the moment and see why he was so bitterly eager for revenge upon Antonio. So he turns to the judge and says, "I stand here on my bond." Here one can hardly fail to perceive or imagine the difference between the first and second utterance of the same phrase, and also between this phrase and the line from Hamlet's soliloquy. Now so great a change could not be caused except by a great increase in mental intensity, by a marked alteration in mental attitude. Obviously the change is not in the words, nor is the essence of it in the altered vocal mechanism (for there is undoubtedly a change there), but it is in the mental concept or picture, in the state of mind, in the feeling which animates the words of the speaker and gives them force and power. Furthermore, the listeners are aware of the change, and that fact is utilized by the speaker.

Let us now consider a case where the situation is reversed.

As I sat at the table the other day I asked my wife, who is also a physician, a very complicated physiological question. She hesitated before answering, then proffered this shrewd reply, "That all depends upon the context." She had hardly time to take another breath when my little three-year-old piped up and said "context." She had caught the closing word and repeated it as she heard it. It is surely fair to suppose that the child in using the word for the first time could scarcely have the slightest idea of the concept that goes with it, of what is meant by the word. How completely then it lost its power and how surprisingly it fell short of any intellectual mark.

This last illustration seems to me the best I could give to make my point absolutely clear. The reason is that we have a word uttered by a mind which is able to clothe the word with meaning and to arouse therefore an idea in another mind. The word is then repeated by a third person in such a way that all meaning leaves it and it becomes a mere sound echoed for the pleasure that the mere sound gives. It shows how blank and empty of concept a word can be, how devoid of weight, thought, or implication. These three examples should make clear how the same words can be so uttered as to give totally different impressions, and how these impressions depend upon the attitudes which mold the mind of the debater. Let us turn therefore to a consideration of the principles that govern these mental attitudes.

THE PRINCIPLE OF MENTAL ATTITUDES

Mental attitudes, or states of mind, take numerous forms, and may be divided accordingly. The usual division is that of thought, feeling, and will, or, more strictly in accord with psychological usage, intellect, affection, and will. In considering mental attitudes under these three headings, however, I do not mean to imply that here we have a cut-and-dried, sharply defined boundary between any two. As in psychology, the division is rather a matter of convenience than of absolute separation, and though they are considered apart, two or even three may occur in any single speech, with one dominating the others. Which dominates at a given time depends wholly upon the subject, the effect one has in mind, and the countermoves of his opponent. For convenience we will consider them separately, and at first, at least, in a general way.

1. *Intellect.* Intellectual attitudes take account of many mental processes. They have chiefly to do, however, with statements of fact, with thinking in the ordinary sense, with con-

sideration and reflection, and with innumerable combinations of these same things.

In the mention of bare facts there is merely the mental attitude of unvarnished repetition. This is the simplest and easiest understood of all the mental attitudes. A statement of fact has little behind it except the speaker's determination to make a certain fact absolutely clear to his hearer. The weight and the impression which it gives is simply the weight of its own truth, and the truth has only the effect of inescapable fact itself. To have any effect, facts must be clearly stated and perfectly understood, but beyond clearness they need ask no more. If, for instance, I say "The railroad curves around at this point in a very sharp, abrupt turn and then continues at right angles," the statement is one of pure thought. It is the maintaining of a truth, the facts of which are undisputed, and which has the weight of truth and no more.

Such a statement of fact is manifestly very simple. *Thought* in general is far more complex. Instead of being confined to a single statement it continues to dominate over a series of them, and it goes from a bare statement of fact such as I have already given to the speaker's interpretation of those facts. This interpretation is really the statement of some conclusion to which the speaker has been led by a consideration of the facts themselves. It is affected therefore by the speaker's attitude toward the facts, and it receives credence in an audience in accordance with the degree of influence which the speaker possesses in that audience. If I am President Eliot, my thought probably carries weight; if I am the Kaiser my authority carries much weight, but my thought carries little. As I am, my thought on some subjects has weight, in others none. Thus in an audience the weight of one's opinion is a variable quantity that cannot always be depended upon to make a point. For example, if I say to an audience "It is my idea that the Allies will win," and if the hearers know that I have had no military training, my opinion falls to the ground. Suppose, on the other hand, I say to an audience, "I think his speech defect can be cured in three months." If I have had a better training and a wider experience in speech defects than any other man in this country, my opinion would be received with respect. This is enough to show that a mental attitude may depend for its effectiveness on circumstances outside the speaker's control and that to give expression to it may be dangerous.

One other kind of thought may very well be employed in debate. That is *logical thought*. Logical thinking is not mere

opinion, mere assertion of one's own ideas, but a careful advance by reasoning from one conclusion to another. The audience is led along the same path of thought and brought irresistibly to the same point that the speaker has reached. This is one of our regular debating devices, of course, and needs no further development here.

Consideration is very like thought in nearly all respects. Yet it differs enough for me to call attention to it. As a mental attitude consideration is the same as "mulling a thing over in one's mind." The Bible says, "Consider the lilies of the field, they toil not, neither do they spin, and yet Solomon in all his glory was not arrayed like one of these." Now a remark of that sort is thrown out to be mulled over, as food for thought, so to say. Ultimately, it will be worked up into a conclusion which will move the audience from a previously held opinion. But that final use does not yet come into one's consideration. We are now merely *considering;* as it were digesting the mental food. As a mental attitude consideration is of tremendous value to the speaker. It elicits the attention of his audience, turns its attention to new things, and imperceptibly moves them from a previous position by the simple request to look at a new one. Combined with the point to which it leads and the final step which it induces, consideration is much like the ordinary advertisement. It has been found by psychologists that the ideal advertisement, the one that attracts a reader, holds his attention, and finally leads him to buy, always begins by telling him to consider something of common interest.

So far as the intellectual process is concerned, *reflection* is much like consideration. More exactly, it is consideration for a second time, a looking back, a re-view, a re-consideration. When you assume a reflective attitude of mind and ask your audience to reconsider data already examined, you appeal to their sense of fairness, and you get a cordial hearing because you appeal to this sense. So if you say to them, "Gentlemen, I beg you to reconsider these facts, in a little different light," you secure a new hearing and a new chance for victory.

2. *Affection.* Under the heading of affection we may consider such things as pleasant and upleasant feelings, the being drawn to a matter under consideration or the being repulsed from it. We have there favorable and unfavorable affections. Shylock's desire for revenge, one's love of his country, and the whole range of personal likes and dislikes illustrate the feelings that we have for people and things. May we ask at this point, what of it? Why consider feeling at all? Why do we have the mental attitudes of like and dislike to-

wards any subject-matter? Why not be intellectual and matter-of-fact individuals, devoting our lives merely to enumeration? The debater's reason is this: He wants to make our audience like and dislike. The thing that he wants the audience to like is his side of the question; in fact, he wants so completely to charm them with his interpretation and his presentation that they will like it to the complete exclusion of the other side.

On the other hand, we want that same audience to dislike the opposite side of the question, and we want them to dislike it to such an extent that nobody can ever say anything that will make them like it again. If I am on the affirmative, I am going to try to make my audience like the affirmative and dislike the negative; if I am on the negative, I am going to make my audience charmed with the negative and disgusted with the affirmative. This I hope to do through maintaining an undercurrent of feeling towards my subject, assured that my audience follows through imitation the feelings of the speaker. It may be important at this juncture to explain briefly the reason for this. It is a large subject, and deserves fuller treatment, but I think that a short illustration or two, or, better, reference to the illustrations that have been given before will lead the reader to agree with me that the audience tends to follow the feeling of the speaker.

The first time that I recited Yankee Doodle, the hearer probably had absolutely no feeling in the matter whatever, but the second time, after I had filled my mind with the conception of the funeral and the sadness caused by the death of the boy's mother, he surely did catch the feeling; and notice, he caught that feeling before I had given him any interpretation of the reason why that feeling should be there. That is, he felt the sadness of it without knowing why I had a reason to be sad. This illustration will I believe prove that the audience tends to perceive and reproduce the feeling of the speaker.

Let me pursue the subject a little further. Take for example the experiences through which Shylock passed before the trial. They give cause for the growth of the Jew's intense desire for revenge upon Antonio. Now it is at this point profitable to ask, what is the effect of that intense desire for revenge upon the audience before which he is speaking? It may be one of two. The audience may keep pace with Shylock and completely sympathize with him, as they probably will if he has carried them along naturally to his final climax; or it may do entirely the opposite; it may take sides with the

opposition to Shylock, and agree entirely with those who plead for mercy. However, as I have watched audiences at this play, feeling has always wavered; that is, the audience sometimes has followed Shylock and sometimes followed the opposition, usually adhering to Shylock if he is not too severe, and then going over to the opposition if it is not too weak in its appeal for mercy. All this makes "The Merchant of Venice" an extraordinary example of the very thing I am trying to illustrate, for it shows how an audience tends to follow the feelings of the speaker, inclining toward revenge or mercy as either is forced upon them by the actor.

Let this suffice for our consideration of the mental attitude of feeling. It is a difficult thing to attempt before an audience, for so to affect people that they will feel kindly toward one's own side of the debate and unfriendly toward the opposite is a subtle and delicate mental performance. Yet difficult as it is, it is worth while. Usually it means victory if one can so accomplish it at the outset that throughout the rest of the whole debate one's hearers will maintain a liking for one's own side and a dislike for the other. I know that oratory is commonly scorned by the academic trainers of speech, but this is not because real oratory deserves scorn. It is because the academician fails to understand the depth to which the non-academic oratorical principles penetrate. I am fitted to talk upon this subject, because I have been trained by both schools, by the academic and the non-academic alike. I am not at all, to be sure, a typical product of either, for I am absolutely out of practice. I can, however, as a teacher bring to others results to which I cannot myself attain. My study in both of these fields leads me to the following understanding of the situation.

The academic school has emphasized throughout in its method of education and training what I call the external method; that is, the word form, the external enunciation, the intellectual cast of the language itself, with program of argument and tabulation of statements and points. On these things they lay the chief emphasis. The non-academic school, on the other hand, proceeds from within outward. They make their chief aim the development of thought processes, the development of power to visualize and picture ideas and conceptions and to throw them clearly upon the screen of the mind. To do this they must of course have a means of expression; they must have speech to tell it all, and words to describe. This they do have, but by no means do they put their chief emphasis on words. Be it far from me to condemn

either of these schools. They both have their value and charm; and they both have these to such an extent that the whole matter of speech in debate is seriously wanting where it falls short of possessing all the excellences of both.

I am also aware of the fact that each of these schools criticises the other unfavorably. The academic school, or the school of the external word, criticises the other because it lacks good English, lacks expressiveness in words, lacks power to put in the word itself all shades of meaning; because it demonstrates too much feeling in its utterances, ("gush," as they call it) and so on. The non-academic school, or the school of the internal thought, criticises the academic because it puts too much reliance upon the ability of words to carry ideas; because it uses no other methods of external expression than language; because it fails to utter the same series of words now in a simple fashion, now with the atmosphere of melancholy interpretation, and now with the blue sky of a sunny disposition, changing the view completely. The non-academic school also finds fault because the other fails to give in its renderings much else besides the word; because it fails to give shades of thought; because it fails to give minute coloring of interpretation; because it fails to give an overshadowing feeling, a tremendous madness, or the utterance of an insatiable revenge, no matter what the words may be. My study in both of these schools has taught me that the non-academic school has by far the best of the argument. I grant that the non-academic school lacks in vocabulary, fails in the presentation of completely polished English, and may sometimes be guilty of the seemingly unpardonable rhetorical sin of using words not strictly according to the dictionary rules; but I maintain that the excellences of the non-academic school can outnumber and outweigh all these little, minor, and pardonable faults.

Some of the things that the non-academic school can do which the academic school cannot do are these: The non-academic school can maintain an extreme variety of mental attitudes over and above any words that one can give them to say. Those mental attitudes they can vary according to the state of mind into which they choose to put their audience, with all the shades of feeling, intellect, and will. They have a tremendous power and sweep of the imagination, and they have their creative imagination so developed that they can picture vast conceptions on the spot. And they can do not only these things; beside all this and to a degree that I have never seen in any academic debater, they can read their audience

and judge from its action what sort of an effect they are making, and from this reading shift their arguments, and move their debating procedures accordingly. Not only can they read their audiences, but they can comprehend the mental attitudes in the speakers of the opposition to a remarkable degree. In fact, they can look through and above the words the man utters and understand his intellectual coloring, understand his imaginative interpretations, understand his shade of feeling in the way that no purely academically trained debater can possibly do.

In brief, the academic people fail to appreciate the introduction of feeling into debates. They believe that usually debate should be without feeling, and, as far as possible, entirely unemotional. The other school insists that emotion has its place, its service, and is one of the most valuable elements that can be introduced in debate if properly and opportunely employed. From all this I may draw the conclusion that the mental attitude of emotion has an unusual and subtle power in debate and deserves serious consideration and cultivation.

3. *Will.* Will is psychologically choice, and the employment of will as a mental attitude means an appeal to the will or an appeal to the choice of a hearer. You know that you can persuade him; you know that sometimes you can drive him; you know that sometimes through feeling you can so elicit his sympathy that he will be led entirely by it. But just here we have a definite requirement. You must lead him to his own choice. That is, you must lead him to what he thinks is his own choice, but really what you, by your skilful presentation, have led him to choose. He must not know of your leadership; he must only know that you present facts, interpretations, likes and dislikes in such a way that he seems himself to choose; and of course you wish your hearer to choose the side that is favorable to yourself.

To lead another to choose is a very complex and difficult matter, but it is an object that one must try to attain all the time one is arguing. If the end and aim of one's speech is to appeal to the hearer to make his choice one must make clear statements of fact, must persuade as far as one can to a point of decision, must reform the hearer's consideration of one's subject, must guide and prejudice his reflection in such a way that he cannot be drawn to later interpretations of an opposite nature, one must appeal to his feeling of patriotism, to his love of home, to his honor, his sense of justice, and his hatred of every thing that is wrong, in order to force his decision and force it to one's side. Then, lastly, the speaker by his mental

attitude of will must force his hearer to make a final change from his own preliminary choice in the matter, to the interpretation presented by the speaker.

To sum up, the mere words of the speaker are the smallest part of the ideas, concepts, and feelings, which his mind seeks to present. Behind the words he possesses knowledge, makes reflections, maintains mental states, develops mental attitudes, and expresses all of these more or less independently of the words, and reproduces them, as far as the hearer is sensitive enough to perceive, in their original form within the minds of the audience.

A CASE OF INTELLECTUAL DEVELOPMENT DESPITE ENFORCED SECLUSION

By JOSEPHINE CURTIS FOSTER, Chief Psychologist, Psychopathic Hospital, Boston

The psychologist who gives mental tests is frequently met by the statement "This low mental age means nothing because the child has been sick and has not been a regular attendant at school" or by "You can't expect a child who has lived on a lonely farm and who has had no playmates to test at age." In vain may the psychologist answer "I admit that inability to read may lower the mental age with a scale like the Binet or the Point Scale. In other respects, however, school training has little influence on the outcome of psychological examinations." Such a reply usually fails to carry conviction to one who is not an experienced examiner. Instead, an objector is likely to depart, thinking that the psychologist is grossly exaggerating the importance of the tests, resolving to take mental ages in the future with as large a grain of salt as ever, and continuing to consider the mental age only when it agrees with his personal opinion of the case.

In our work at this hospital we have come to use the answer "Let me tell you about a fifteen year old boy we examined some months ago." The illustration has been so successful that we offer it here to other psychologists for similar use.

The Story of John

John Henry[1] was born in a rather poor district of a New England city on August 13, 1902. Of his remote ancestors we know comparatively little. Presumably his paternal grandparents were natives of Ireland. His maternal grandparents are said by one of our informants to have come from England and Scotland, and by another to have been born in Nova Scotia and Maine. We have one history which says that a maternal uncle died in an insane asylum. Another maternal uncle, who came to the hospital upon request, is reported to have been insane, appeared to be very unintelligent and was considered

[1] This, of course, is not the real name, but it is like it in having the same second name (which the boy spelled Henery).

as "probably untrustworthy in his statements." This uncle denied mental illnesses in the family, but at the same time showed such familiarity with hospitals for the insane as to suggest personal acquaintance. Of the immediate family, the boy's father was born in Ireland, and was probably a horse-trainer. He was heavily alcoholic, abused the children, and some five years ago deserted the family. John's mother was born in New England. She was at this hospital for a time, and is now an inmate of a state hospital for the insane with a diagnosis of Dementia Praecox, paranoid. Four of John's siblings are said to have died of starvation in the first year. An older half-sister is apparently normal.

The schooling of John and this half-sister, whom we shall call Julia, was decidedly meagre. They were allowed to attend occasionally, but although both of them learned to read, they acquired little else. If we could add together all the days John spent in school, the total might come to three years. We know that the family moved so frequently (at least eleven times in four years) that he could seldom have spent a full term in any one school. John was last in the fourth grade. He attended school in this grade irregularly for some years. Apparently he was not held from advancement on account of dulness, but because the mother went to the teacher and said "Don't waste any time over this boy of mine. He is feeble-minded and will never learn, no matter how much effort you expend on him." Apparently, also, the teacher believed the mother who was so candid about the failings of her son, accepted her advice, and spent her energies on the other children.

As might be expected under the guidance of such a mother, home life was not an aid to John's mental development. Julia says that her earliest remembrances are of squalid surroundings, a room or two with the poorest of furnishings, which were never anything but filthy. She had very little clothing, and at times such a small amount that she was forced to stay indoors. John says that his mother would buy him new clothes and then lock them up till he had outgrown them. Moreover, there was never enough food, and the children were continually hungry. The mother and step-father beat Julia and obliged her to chop wood and do all the work. In her presence they also indulged in perverse sexual practices. John was not abused to the same extent, probably because he was the father's own child. However, the only attempt to teach him anything was his father's instruction in obscene verses and songs.

John and Julia had not even the opportunity to learn from

contact with other children. Their mother often kept them in the house, and this happened still more frequently after the father deserted. Even when the mother was away scrubbing in the middle of the day, the children dared not venture out, for fear there might be some truth in their mother's statements that "there was a God lover head" who would do various things to them, that the policemen would get them and do terrible things, that people would kill them, and later, when Julia was living elsewhere, the mother told John that "they would catch him in the draft"—a hazy, but horrible fate.

John's story of his life is of a succession of moves from house after house, with periods of confinement varying from a few days to one that lasted 13 months. Most of the daytime he slept or amused himself by reading the Bible, the few religious books and the few Indian stories which they owned. When the mother returned from her work, he seems to have been occupied with endeavoring to escape her wrath. He says that she would ask him questions and then answer them herself. If he attempted a reply, she usually slapped his face. She talked continually of the references which the newspapers made to her, and of the pictures the moving-picture men were taking of her. They had no regular meals. The mother probably ate something at the restaurant where she did scrubbing, but all she brought home to John was bologna sausage, milk, crackers, and bread. When he came to this hospital, the boy had not had a hot meal in months. At night he was often kept awake by his mother's talking, by what he called "Mother's lies."

Things went from bad to worse with John, but they improved for Julia in the spring of 1917. At that time the mother accepted a position as "an experienced cook" and took the children with her. On the third day she was discharged as dirty and hopelessly incompetent. Julia, however, was retained and since that time has had normal surroundings. John and his mother returned to their squalid living, and the boy was confined even more closely until early in 1918, when the police and one of the children's societies found them and brought them both to this hospital.

When John was examined here, he was a large, overgrown boy, sallow and anemic. His voice was husky and uncertain. He did not know how to meet strangers, and was shy and reserved. His face was a little sober, but almost expressionless. He smiled slightly and then only after many interviews. He answered all questions as briefly as possible. The most surprising thing about his attitude was that he took practically

no interest in his surroundings or in what was to happen to him. John's ignorance of the world about him was astounding. He could name the president, the governor, and several recent mayors, but when asked for the capital of Massachusetts he did not understand what was meant. He knew nothing whatever of geography. With the exception of reading, arithmetic was the only school subject in which his performance was at all creditable. He had never seen a moving picture or a baseball game.

We have here, then, a boy of fifteen and a half, who has had practically no school training, practically no contact with other children, and no home education save what he may have picked up from the reading of a few religious books, and perhaps a stray copy of a newspaper.

In spite of this overwhelming lack of information and of training, when given a psychological examination John received a mental age of 16.0 (I. Q., 1.04) on the Yerkes-Bridges Point Scale, and a mental age of 15 yr. 1 mo. (I. Q., .98) on the Stanford Revision of the Binet. There is nothing in the examinations that would be unusual in a normal boy of his age. On the Point Scale the errors he made were in: repetition of 7 digits, repetition of the longest sentence, failure to give interpretation for two of the pictures, failure to give definitions in terms of classification, acceptance of the line suggestion, failure in one of the memory drawings, in one of the definitions of abstract terms, and in one analogy. On the Stanford scale, he passed all the 9 year tests, failed in designs from memory in year 10, in interpretation of pictures in year 12, in the vocabulary in year 14 and the succeeding years, in giving differences between abstract words, in repeating backwards, and in the code in year 16, and in the repetition of digits forwards and backwards and in giving the thought of a passage in year 18. His arithmetical ability is particularly evident in the Stanford examination. He did the Healy Construction Puzzle A in 7 moves in 15 seconds and Construction Puzzle B in 12 moves in 30 seconds. His attitude during the psychological examination was somewhat listless and a bit childish, but he cooperated well and showed occasional interest and pleasure.

The conclusion must be that, in this case, at least, lack of school-education, lack of home-training, and lack of normal intercourse with other children have not prevented a normal rating on intelligence scales.

John's subsequent history, while not of consequence for our purpose, is, nevertheless, interesting. He was discharged from this hospital

to a society for placing and caring for children, with the diagnosis "Not Insane, not Feeble-minded." We hoped that it would be possible to give an examination at the expiration of his first year of living in normal surroundings. This proved inadvisable because the boy is doing good work on a farm, and a trip to the city might encourage his present ambition to work in a store, a job which doubtless would not fit him just now. The society, however, sent us a detailed report of a "visit" made and we give the substance of that report here:

John is living on a farm with a family who understand his case and are anxious to assist him in every way possible. He has gained in weight since last year and he looks physically well. However, he is very slow in his movements, and appears to take no interest in his surroundings. He seemed little affected by the fact that the visitor had come from Boston to see him, and the family say that this is his usual attitude. In the year that he has been with them, he has never spoken of his mother and sister, has never asked questions about places to which he has been taken in the automobile, or shown any interest in them. The only exception was one trip to the ocean, when he asked "Why can't you see the other side?" But he has never spoken of the trip since. His typical day is as follows: Rises at 6, without being called, feeds the horses, without being told, milks three cows, comes into the house for breakfast, without being called, eats a hearty breakfast, then without any comment, returns to work, waters the cows, and cleans up the barn. He comes back to the house about meal time, gets the mail from a box at the end of the lane without reading any addresses or ever asking if there is a letter for him. After the noon meal he takes care of the hens, and does odd jobs. In the summer time he goes for the cows. After supper he reads the jokes in the daily paper and Farm Journal, sometimes smiling a little. He never grumbles while doing anything he is asked, never asks a question, very seldom speaks without being spoken to, goes to bed without being told to at 6:30 and is perfectly willing to do this day in and day out. In fact, far from disliking routine, he is somewhat disturbed when rain prevents the performance of some of his usual chores. He shows no interest in anything. He is good to the animals but never pets them and the dog and cat never ask him for food. He never notices anything unusual. If, for example, a cow were in great distress and he had not been told to notice whether or not the cow was sick, he would make no comment. Neither is he at all interested in the fact that the farmer is willing to give him $50 if he remains with him a full year. Besides this total indifference, the farmer and his wife have reported several facts which are of importance in the prognosis of the case. He has recently been smiling more often, and occasionally laughs heartily when alone at his work. When asked what he was amused at, he insists that he did not laugh. He has never made any attempt to play with a boy who lives in the next house, a quarter of a mile away. It seems likely that the boy may end by having Dementia Praecox, like his mother.

BRIEFER STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF McLEAN HOSPITAL

Reported by F. L. WELLS

		PAGE
I.	The "Coachproof" Test	172
II.	A Filing Test	178
III.	A Brief Test for Mental Accuracy	179
IV.	The "Cued Combination" Method	183
V.	Geographical Information in High School and College Students	186
VI.	An "Omnibus" Test for Reasoning and Imagery	189
VII.	Certain Individual Records Compared with Standard File	191

(The experimental work of these studies was complete on July 15, 1917.)

I. THE "COACHPROOF" TEST

1. The "clinic rounder," vitiating Binet findings by previous experience of the tests, early demanded the protection of such measurements against foreknowledge. As mental examinations are introduced in various kinds of personnel selection, and it becomes increasingly to one's interest to make a good score, protection against coaching in the methods used becomes altogether necessary.

2. A test is "coachable" to the extent that superior specific ability in it may be acquired through practise in it, or through knowledge of the result to be attained. Tests of involuntary processes are least subject to coaching. Sensory tests are somewhat coachable, though it is questioned whether the improvement is in strictly sensory acuity, or in more efficient use of accessory cues. Voluntary motor tests are coachable to the extent of practise improvability, which may be considerable if not rapid. With well designed experiments, the practical danger of coaching in these fields is not great. The danger of coaching is greater in the higher mental processes, where right response is determined by factors of conscious knowledge. One simply learns that he is to speak certain words or draw certain lines, in response to particular problems. While such coaching may be unmasked in individual examination, it is none the less vitiating thereto. In group experiments it is less likely of detection.

3. There is little protection against special ability acquired by practise in the actual motor or sensory tests. Protection against coaching in tests of the higher mental processes is provided through alternative tests of similar material, but so differing that the literal responses to one are not correct to

another. Tests with a number of homogeneous alternatives are for practical purposes regarded as commensurable and coachproof.

4. Though essential test elements are changed in the alternatives, they preserve common factors open to mutual practise effect. Such tests are still coachable to the extent to which practise in any one form of the test will improve performance in its alternatives. In this way is measured the degree to which a test is coachproof. This problem is essentially the not unfamiliar one of practise and transfer.

5. In the course of preparing material for coachproof instruments of mental examination, inquiry was made into the perfection with which the method of alternation attained this purpose. Ten alternative forms of the Woodworth-Wells hard directions test were prepared.[1] A subject accomplished one of these and one of the Woodworth-Wells standard blanks daily for ten days. On odd days the standard form preceded, on even days the alternative form. Then ten standard blanks and ten alternative forms were on the same day accomplished in immediate succession. There was thus practise in an identical task, paralleled by practise in a continually changing task. Evidence of coachability in the test is the appearance of practise in performance with the alternatives.

6. The score is in seconds' time. For each error, the time consumed is increased by 1/15 of itself (there are 15 responses in the test.) With this correction for error, the initial time of the standard form is 92 seconds, of the alternative form 94 seconds. Figuring these starting points as 100 per cent, the per cents of each reached in the subsequent tests are shown in the accompanying curves.



[1] For illustrations, cf. *Journ. Appl. Psychol.*, 1917, vol. 1, p. 139.

7. Both standard and alternative tests show significant practise improvement at the end of twenty performances. The best performance in the standard test is 24 per cent of its original score. The best performance in the alternative test is 40 per cent of its original score. The subject's performance at the beginning is about the normal average. After practise periods totalling less than an hour the alternative forms are accomplished in time near the lower limit of the unpractised normal range. A test having this property is not well fitted for use where there is special motive to make a good showing, and normal opportunity to train therefor.

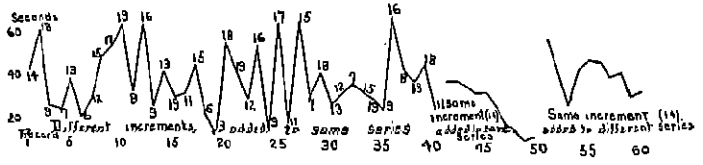
8. Other experiments were made in which the practise was not obtained from day to day but in immediately successive performances. Practise was given in a "constant increment" test, adding 17 to each of a series of 10 two place numbers. Time and errors of each series were noted, the experiment consisting of 45 series. The first three series were different, and average 36 seconds without correction for one error made. The third series was then practised for seven records, without error, the time decreasing to 21 seconds. The subject having then studied for 3 minutes to learn the series, practise was continued to the fifteenth record without change. Then new series (Nos. 4-8) were introduced for five records. They were accomplished without error, in average time of 29 seconds. Then series Nos. 1-8 were repeated (records 21-28). Series 1 and 2 showed about 68 per cent of their beginning rate, the others were longer than their previous records. Record 29, a wholly new series (No. 9) is 28 seconds and errorless, clearly better than initial performance. Practise being resumed on series No. 3, (records 30-35) the subject was requested to memorize this series and report when done, (224 seconds). The errorless time dropped to 14 seconds. Records 36, 37, new series, have 68 per cent of the initial score; 38 is very long, apparently blocked. In the remaining records, 39-45, series 1-7 are used with the addition of 16 instead of 17. Practise improvement begins to show at the close of the experiment. There was subjective feeling of "ready to stop," as though doing slowly.

9. Continuous experiments like this are objectionable as introducing a fatigue (refractory phase) element not present in practical application of the tests. This tends to mask gains by transfer. On the other hand, there is an accustomedness to the whole physical setting of the experiment that is helpful, but absent in the conditions of practical application. A somewhat more satisfactory situation is indicated than in the hard directions test. The best performance with the same incre-

ment in a new series is 74 per cent of initial performance. The best performance in the practised series is 41 per cent of initial performance, representing the time of writing a memorized series. When the increment is varied (adding 16 instead of 17), the gain is practically neutralized.

10. Another subject accomplished 40 records in a constant increment test from a single series, No. 4. In these records different increments were added to the series. These numbers added ranged from 3 to 19. The increment in the initial series was 14, and during these 40 records this number is not so employed again. In records 41-50, the increment 14 was added to series No. 4 in all cases, thus giving practise on this series and number only. In records 51-60, 14 was added to different series, save that the last two records were again with series No. 4.

11. The score is the time consumed for a series, plus 10 per cent thereof for each error. These scores for the 60 records are shown in the accompanying graph. Attaching to each point of the curve for the first 40 records is a special figure. This denotes the figure added in that record. In records 41-60, 14 is always added; to the same series in records 41-50, to different series in records 51-60.



12. The practise in adding other figures to series No. 4 does not appear to increase facility in adding 14 to series No. 4. In records 41-50, the repeated adding of 14 to series No. 4 rapidly improves capacity in that specific function. In records 51-60 it is not evident that the previous 50 records have improved the adding of 14 to any other series than the one used in those 50 records. It results that practise with the same series and the same increment gives great improvement. *Different series with the same increment give some improvement.* The same series with different increments gives very little if any improvement. In practical experiments it is easy to vary both series and increment. It is indicated that superior capacity for an unforeseen series and increment is not to be gained save through a generally increased capacity for all series and all increments. A test with this property is coachproof.

13. Other experiments were performed in which the task is the rearrangement of various items in a serial order. A

subject was instructed to rewrite in order of magnitude a series of ten 4 and 5 place numbers. The first two records were with different series. Series No. 2 was then practised until the tenth record. The 11th record was a new series. Practise was continued with Series No. 2 until the twenty-first record. Nine new series were then introduced, making thirty records in all.

14. The score is the time in seconds for accomplishing a series, plus 20 per cent of itself for each break in the series, plus 10 per cent of itself for each lapse in writing. Fourteen out of the thirty records taken were penalized for error.

15. Records 1 and 2 are errorless and average 46 seconds. The last two errorless practise records, (19 and 21) average 46 per cent of this score. The last two errorless new series, (records 25, 26), average 40 seconds, 87 per cent of the initial score. The final group of new series, (records 22-30) averages 130 per cent of the original score, so heavily were they penalized for errors. No marked gain in the accomplishment of new material is indicated.

16. In a subsequent experiment the procedure was changed in that the subject instead of writing the series itself in order, numbered the items to show the proper order.

17. The score is the time consumed in accomplishing a series, plus 20 per cent for each break in the series. Of the 25 series accomplished, 14 are without error.

18. The initial time is 35 seconds, the score being raised by one error to 42. The responses improve quickly through memorization. The seventh record is 19 per cent of the original score, beyond which no improvement is observed. The best score in a new series is 52 per cent of the original score, (record 17). The best score in a new series penalized for error is 73 per cent of the original (record 21). It is indicated that the practise effects some general improvement; less coachability appears than in the hard directions test and more than in the constant increment test.

19. In another subject, the experiment was performed not with numbers, but with names to be numbered in order of alphabetical sequence. The first two records were with different series. Series No. 2 was then practised until the ninth record. The tenth record was a new series. Practise with series No. 2 was then continued until and including the twentieth record. Ten new series were then introduced, making thirty series in all.

20. The score is the time in seconds for accomplishing a series, plus 20 per cent of itself for each error. Six of the thirty series were thus penalized.

22. The improvement in the practise series is to 15 per cent of the original score. The best score in a new series is 55 per cent of the original score, (record 30). The best score in a series penalized for error is 65 per cent of original score (record 22). Similar coachability to that with the consecutive numbers is indicated.

23. Attention is invited to already reported results on practise with the free association experiment.[2] With no repetition of stimulus words there are general decrease in time, and other improvements of performance. Records of this test may be vitiated where there is motive to attain high degree of practise. There are also certain instructions about responding whose observance would improve the score if not unmasked in lengthened times and the character of responses.

24. The number-checking test [3] employed in connection with the above association experiments is not alternative and thus without significance for the present question. The addition material (Kræpelin's *Rechenhefte*) is alternative and shows considerable practise. It is not probable that the subjects' general quickness at figures is increased so greatly if at all. For quickness at figures, such a test is invalidated in special practise, the increased capacity being confined to a particular and practically unusual sort of adding.

25. Such a test of adding is very analytical; it does not cover the field of quickness at figures, but a specific part of it which will normally be an index of the whole. If a subject is specially practised, this relation is distorted. As an analytic test, a test is "coachable" in any form in which it shows practise improvement.

26. Alternative "typewriting tests" are coachproof in spite of their practise improvement, because such improvement applies more equally over the general accomplishment of typewriting. Instead of analysis, the typewriting test aims at synthesis, that is, at reproducing in miniature as much as is possible of the practical accomplishment. Where, as in typewriting, a reasonably synthetic test can be made, "new" test material is also coachproof.

27. The practical significance of mental tests, except the specifically "trade" tests, is analytical. It is indicated that alternation leaves certain tests open to not inconsiderable practise improvement. Such tests can still be "beaten" where the practical motive is sufficiently strong.

F. L. WELLS.

[2] Practice Effects in Free Association, *Am Jour. Psych.*, 1911, vol. 22, 1-13.

[3] Relation of Practice to Individual Differences, *Am. Jour. Psych.*, 1912, vol. 23, 75-88.

II. A FILING TEST

28. The essential operations of filing are searching the files for a given entry, and filing a new or removed entry in the file. Efficiency of file clerks is given in the speed and accuracy with which these operations are accomplished.

29. The part played by memory, patience, and general intelligence in this accomplishment, increases with the complexity of the files to be handled. The filing test proper is concerned with specific aptitude in the psychomotor processes of manipulating the file. The task is to find the place where a given entry is or ought to be.

30. A testing file of 500 cards, 2½ by 3 inches, contained in a standard file drawer, was furnished for the present purpose by the Phœnix Mutual Life Insurance Company.

31. For the test of searching fifteen cards at random were removed from the file, and five other names in it were noted. All twenty names were written on a single sheet, which was handed to the subject, with instructions to find whether the names were present in the file or not. If the name was found, this was announced by speaking the name; if not, the announcement of the name was preceded by the word "No." Another procedure was to have the names writen one each on memorandum slips and have the subject verify them as above. This made it possible to present the same names in varying order. Each operation was timed with a stopwatch.

32. For the test of "filing" proper, twenty cards at random were removed from the file. They were given to the subject with the direction "File these cards in their right positions. Let them rest simply on the rod" (the locking rod as shown) "without removing the rod and pushing them down into place." The time of filing each successive name was noted. Afterwards the subject might be asked to check the correctness of the cards just filed, a process taking from 50 to 170 seconds. The correctness of the filing was verified before removing the locking rod and returning the cards to their final positions.

33. Results from eight subjects are as follows:

Average time in Seconds in Elementary Filing Operations.

Subject	A	B	C	D	E	F	G	H	Av.	ratio of M. V.
Searching	15	21	25	19	35	25	19	18	22.1	.21
Filing	25	17	14	19	37	26	17	15	21.2	.28

34. The first five subjects are without special practise in office methods; the last three are office secretaries, but not professional file clerks. There is evidence that such times as

the above are much reduced in special training. There is general correspondence in the speeds of searching and filing, which is confirmed in office experience. Subjects A and C however, are quick in one operation and slow in the other.

35. Final scores are stated in terms of time. Errors do not essentially affect the present results. In general experimental practise, it is best to correct a time score for errors through increasing it by a percentage of itself. Heavy penalty is to be imposed for errors in a filing test, such as ten per cent of the time score for each per cent of error.

36. File manipulation is influenced by external factors such as the number of names in the file and their dispersion through the alphabet. The names in the present file included but a small fraction of the letter A. A file with the same number of names but running throughout the alphabet, would probably be easier of manipulation.

37. The above represents a standardizable form of test for filing operations proper. It is coachproof. It was in the present instance applied as an individual test, but is convertible into a group test with sufficiently extensive testing files.

F. L. WELLS.

III. A BRIEF TEST FOR MENTAL ACCURACY

38. It was desired to arrange series of tests which should be (1) coachproof, (2) suitable for application to groups, (3) applicable to individuals with less demand upon the experimenter's attention than time-limit group tests. This and the following studies deal with tests so arranged.[1] In practical application these tests are most suited for the selection of clerical workers.

1. Add 17 to each of these numbers. Write each answer on the dotted line to the right of each number; thus 10 27.		*2. Divide*	*3. Rewrite this list of numbers, putting the largest number on the dotted line at the top, and the others in order of size below it.*
34.....	31.....	384) 74496	8846 .....
79.....	48.....		9059 .....
88.....	66.....	761) 195577	10604 .....
68.....	26.....		10653 .....
25.....	60.....		10097 .....
82.....	98.....		10578 .....
27.....	37.....		10585 .....
30.....	13.....		8605 .....
24.....	52.....		7895 .....
23.....	78.....		8813 .....

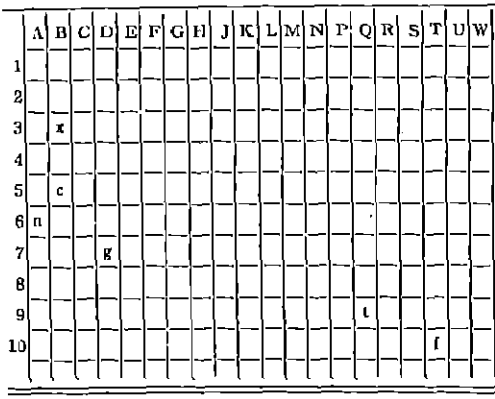
[1] A brief time limit intelligence test was also devised, described by G. G. Fernald. Report of the Maine Commission for the Feebleminded, 1918, pp. 11, 75.

4. *On the dotted line before the name coming first in alphabetical order, write the figure* 1; *before the name next in alphabetical order write the figure* 2; *and so on until you write* 10 *before the name last in alphabetical order.*

..... Foley, Annie E.
..... Gates, Mary M.
..... Kennedy, Allan T.
..... Mac Donald, Alice J.
..... Robertson, Clara L.
..... MacDougall, Vincent
..... Judge, Mary
..... Mingo, Ida L.
..... Marden, Edith
..... Roberts, Fannie

5. *Copy the numbers given below, putting them on the dotted lines at the right in the same order as they are given.*

5148G3729 . . . . . .
169472853 . . . . . .
3841259617 . . . . . .
837594162 . . . . . .
392617584 . . . . . .
471862953 . . . . . .



6. *Each of the squares in the diagram at the left is named according to the letter under which it stands and the number on the same line with it. Thus the square with an* x *in it is named* B3.

Give the letter and number to name the squares which have in them the small letters c....; f....; t....; g....; n.....

Put a figure 6 *in square* W2
" " 8 " " N4
" " 2 " " F3
" " 7 " " G4
" " 4 " " J7

7. *Do what it says to do and be sure you get just what it says.*

If New Year's comes in May, make a figure 6 here....; but if not tell where the sun rises..... Give a wrong answer to this question, 'How many days are there in September?'.... Write "no" no matter whether Canada is in Africa or not..... Write the second letter of your first name and the first letter of your last name at the beginning of the dotted line..... Notice the numbers 2, 7—If iron is heavier than water, write the smaller number here....; but if iron is lighter, write the larger number here.... If people believe that Columbus was King of England, cross out what you last wrote, but if not, put in the number to complete this sentence, 'A cat has .... ears.' If Friday comes after Thursday, make two crosses here.....; but if not, make a square here.....or a circle here..... Give the correct answer, "yes" or "no," to this question, 'Do apples have four corners?'..... Write again what you last wrote, here..... Make a comma after the shortest of these three words, hat pill ladder. Put a square before this name of a boy.....John. Write any letter except "e" just after this comma, ... Make a dot to the right of any one of these letters F G H I J. Show by a cross when the days are shorter: in winter?.... in summer?.... Write "no" if 2 x 6 are 12.....; if not, make a circle here.....

39. The present study concerns the first of these test-series, termed C. A form of it is reproduced herewith. It is seen to be capable of indefinite alternation. Actually ten forms in

addition to the above have been printed, omitting however test 7, the alternative of the Woodworth-Wells Hard Directions test.

40. It is intended that the problems of this test-series shall all be well within the range of the subject's information and experience. The quality of performance then depends in first instance upon the *care* with which the tests are accomplished, not upon special information of correct answers or ability to reason them out. In subjects for whom the series is suited, performance is limited essentially by the subject's carefulness. Other test-series are limited by knowledge and by reasoning ability.

41. For accuracy the tests are scored as a point scale. The sum of possible credits is 100. Each item represents a proportional part of these credits. Deductions therefrom are made for errors uncorrected. In tests 3 and 4, a deduction is made for corrected errors, which is doubled for uncorrected.

42. A rational final score in tests of this nature, or any function thereof, is the per cent of all records in file which the record in question surpasses. If it is the best of all records, it thus scores 100 per cent; if better than half, only 50 per cent, etc. A card file of results is kept, in which each new accession to the data is entered in order of its amount of credit.

43. The present data concern 140 individuals; 75 records from college students, 53 from high school students,[4] the remainder from adults of the same order of intellectual capacity as these. From the college group, the per cents surpassing different scores in points earned, and in minutes of work, are as follows:

Point Score	Per cent surpassed by same	Time for test (min.)	Per cent surpassed by same
20	0	22	0
30	1	20	1
35	3	19	4
45	4	18	8
50	11	17	13
55	17	16	20
60	24	15	28
65	29	14	38
70	40	13	53
75	52	12	67
80	61	11	82
85	73	10	94
90	84	9	98
95	90	8	99
100	97		

[4] For this material the writers are indebted to the courtesy of Dr. H. T. Woolley.

44. With the high school group, the time score was in terms of points earned per minute of work. In this group, percentile values of various scores are approximately as follows:

Point Score	Per cent surpassed by same	Points earned per minute	Per cent surpassed by same
31	0	1.0	0
40	8	1.5	4
50	15	2.0	4
55	19	2.5	7
60	32	3.0	19
65	40	3.5	32
70	47	4.0	45
75	54	4.5	60
80	60	5.0	68
85	74	5.5	79
90	83	6.0	87
95	93	6.5	91
100	96	7.0	95
		7.5	98

45. When graphs of scores for both groups are superposed, there is evident no separation of the groups in respect to number of points earned. The ranges of performance are similar in extent. The number of points earned per minute is clearly greater in the college group. Their superior mental discipline enables them to perform the operations faster, but they are accomplished with equal accuracy in both groups.

46. The accompanying table analyses the error penalties in different tests of the series.

PER CENT MAKING

	No errors	1 error	2 errors	3 or more errors
1. Addition	48	25	10	17
2. Long Division	73	21	6	..
3. Consecutive numbers	44	21	15	20
4. Alphabetical names	50	14	8	28
5. Copying	93	4	3	..
6. Letter-square	60	19	3	18
7. Directions	27	36	21	16

47. As is consistent with paragraph 45, the different items are of nearly equal difficulty for the high school and college groups.

C. M. KELLEY,
F. L. WELLS.

IV. THE "CUED COMBINATION" METHOD

48. The present study deals with performance of high school and college students as indicated in paragraph 43. The test-series is termed K, the form used being reproduced on the page opposite. It employs a variant already described[5] of the combination method, and is a preliminary form of the test-series, to which similar alternatives have been prepared.

49. The first portion consists of words 35-70 inclusive of the Terman Vocabulary Test, omitting the words *stave, snip,* and *swaddle.* The cues in this portion are intended to be such that the word cannot be supplied unless known. The second portion contains words of special difficulty in spelling. The cues for these words are intended to indicate the word clearly, without indicating its correct spelling. The third portion presents another kind of material, and is discussed subsequently.

50. This test-series differs from series C in that it is limited by the subject's absolute knowledge of correct responses rather than by his carefulness in setting them down. Carelessness is less likely to vitiate what is known here than in series C, because there are more logical connections by which responses check themselves up. Carefulness adequate to perfect score in series C avails nothing for the responses of series K outside the range of knowledge. Practically, this type of test-series is indicated where clerical duties take on a secretarial nature, more intellectual than reached by the mechanical tests of series C.

1. *Each of the sentences below has a word of which only certain letters are given. You are to fill in whatever letters are missing so that the word is complete. If you are not sure you have filled it in rightly, you may put a ? after your answer. If you cannot fill in a word in about 15 seconds, go on to the next.*

Each period (.) in a word means one letter left out; a long dash (———) means any number of letters left out.

Examples:

The boy was sucking a big ripe or...e (orange).
A cent is made of co———(copper).

This policy will .ns... your house against fire. We r.g..d your idea very favorably. The dentist was obliged to kill the n.... of the tooth. We heard the carriage-wheels c..nch the gravel. The j.....r kept six balls going in the air at once. The judges in their robes represented the m...sty of the law. She is of a deep brun.... complexion. He copied his superiors in an a..sh manner. The pup is a sp..ti.., frolicking little animal. She laughed herself into h.....cs. M... was

[5] *Journ. Appl. Psychol.,* 1917, Vol. 1, p. 140. *Journ. Educ. Psychol.,* 1917, Vol. 8, pp. 483-487.

the Roman war-god and is the name of a star. He left to seek r.p..e from his labor. You are too shr... to be tricked by such talk. Who loses the game must pay the f..f..t. He has a marked p.c.....ty of speech. The c....ge of 2½ cent pieces has been thought of by the government. The table was inlaid with beautiful mos... designs. They ceased to bew... their misfortunes. The labor of doing that is disp.....nate to the good it does. The house, old and d...p....d, is still an interesting ruin. The company was granted a ch....r by the state. He is very con....n..ous about always doing his best work. The greatest fault of the miser is av....e. The story was told with ar....s childlike simplicity. These records are pr....ess, and must be kept with greatest care. We will t....ate these abuses no longer. It was a g...t.nous mixture, soft and elastic. The North Cape is a high p.....ty extending into the Arctic Ocean. Food in camp must be guarded against the d....d.tions of squirrels. His designs have been learned in time to f..str... them. He is a weakling, a m..ks.p afraid of all hard work. He was devoted to ph.....py, and gave much to his home town. He spoke with bitter i...y of the abuses he had seen.

This work is the crowning a.....ment of his career. The shipment is due day after t————. This article is ind.sp....ble to all householders. We cannot allow a further cont....nce of this practise. You will super.e.. him in the control of the job. Be the first to s.... these exceptional opportunities. Several men pre————ded him in this observation. A musty odor will ———— nate from disused cellars. It is our b————ness to see that the work is well done. The letter was not at hand because the boy had f————ten to mail it. To su———— in life one must apply himself steadily. We should be guided by the same prin———— in all these cases. She met her fian———— at the railway station. We have received mon————s amounting to over $7000. This is a sad oc————nce, which will not happen again. The two sep———— firms will unite to form one. We are ref———— to you by your friend, Mr. John Robinson. This process will ——im.nate waste and result in great saving.

2. *Below are names of towns and cities in the United States. After each name, write the name (or abbreviation) of the State in which that town is. If the same name is given more than once, give a different state for each time it is given. Indicate only the larger of the towns so called for, not smaller ones like Salem, Maine, or Rutland, Mass.*

Skip any you do not know, and go on to the next.

Muskegon.....	Duquesne.....	Alexandria.....	Camden.....
Ansonia.....	Newton.....	South Bend.....	Rock Island.....
Walla Walla.....	Huntington.....	Paterson.....	Chattanooga.....
Hagerstown.....	Boisé.....	Jackson.....	Malden.....
Leominster.....	Washington.....	Jackson.....	Pottsville.....
Cambridge.....	Washington.....	Jackson.....	Belleville.....
Aurora.....	Erie.....	San Francisco...	Oswego.....
New Castle.....	Berkeley.....	Norwalk.....	Kenosha.....
	Topeka.....	Michigan City...	

51. For accuracy, the test-series is scored as a point scale, the sum of possible credits being 100. Each item represents a proportional part of these credits. Records were kept as

indicated for Series C in paragraphs 41 and 42. The question mark feature of the directions was without value.

52. Percentile scores for the 75 college students and 53 high school students, for the entire test-series, are as follows:

College Group				High School Group			
Point score	Per cent surpassed by same	Time of test (min.)	Per cent surpassed by same	Point score	Per cent surpassed by same	Points earned per min.	Per cent surpassed by same
41	0	26	0	31	0	1.0	0
45	3	22	9	35	9	1.5	8
50	8	20	19	40	25	2.0	28
55	23	18	43	45	45	2.5	53
60	44	16	61	50	59	3.0	73
65	65	14	76	55	79	3.5	83
70	82	12	92	60	93	4.0	91
75	94	11	96	65	98	4.5	96
82	98	7	98			5.0	96
						5.5	98

53. There is more separation of the college from the high school group in this test-series than in series C. The present test-series is of a type calculated to specially bring out the formal educational difference involved. The separation is somewhat greater in respect to total points earned than in points earned per minute.

54. The accompanying table analyses the credits earned in the items of the cued combination test, all data being combined.

Word	insure	regard	nerve	crunch	juggler	majesty	brunette	apish	sportive	hysterics	Mars	repose	shrewd	forfeit	peculiarity	coinage	mosaic
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Correct, %.....	91	98	98	68	68	52	85	37	18	67	91	56	80	86	59	81	68
Failure of spelling, %.....	3	0	0	4	21	10	11	1	1	25	2	1	11	8	6	2	1
Wrong word combined, %...	5	0	1	20	8	16	1	15	36	3	2	7	3	2	1	9	20
Omitted, %.....	1	2	1	8	3	22	3	47	45	5	5	36	6	4	24	8	11

Word	bewail	disproportionate	dilapidated	charter	conscientious	avarice	artless	priceless	tolerate	gelatinous	promontory	depredation	frustrate	milksop	philanthropy	irony	achievement
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Correct, %.....	73	47	23	90	43	53	23	58	79	25	21	22	67	29	61	79	29
Failure of spelling, %.....	0	14	51	1	37	12	1	1	1	8	36	11	7	1	6	3	3
Wrong word combined, %...	7	3	2	1	1	8	14	7	4	21	8	6	4	0	3	0	17
Omitted, %.....	20	36	24	9	19	27	62	34	16	46	35	61	22	70	30	18	51

Word	tomorrow	indispensable	continuance	supersede	seize	preceded	business	forgotten	succeed	principle	fiance	moneys	occurrence	separate	referred	eliminate
	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Correct, %.....	88	44	80	32	30	35	94	82	80	71	69	45	49	91	60	48
Failure of spelling, %.....	1	52	17	36	15	31	6	8	11	4	25	22	48	7	22	13
Wrong word combined, %...	6	2	1	8	29	13	0	2	5	24	6	7	0	0	17	9
Omitted, %.....	5	2	2	24	26	21	0	8	4	1	0	26	3	2	1	30

55. Special disparity, not shown with the tables, between college and high school groups is indicated in the words *apish, gelatinous, depredation* (*s*), *achievement, seize*.

C. M. KELLEY,
F. L. WELLS.

V. GEOGRAPHICAL INFORMATION IN HIGH SCHOOL AND COLLEGE STUDENTS.

56. A list was made of towns and cities in the United States having population of over 15,000 by the 1910 census. The names were arranged in random order. The list includes 400 names. Ten alternative lists of 34 names each were prepared. One of these is the basis of the present observations, its material being reproduced with paragraph 48.

57. Responses were not scored correct unless they designated a town included in the population requirement, irrespective of there being elsewhere a smaller town as named; (cf. Salem, Maine; Rutland, Mass.; in instructions.) As a test of information, it is simply of whether the larger town in question is known; further, it is not evident whether the designations of smaller existing towns represent actual knowledge thereof, or have only chance correctness. Credit is allowable however, for such responses as Lakewood, New Jersey or Plymouth, Mass., where larger but less widely known towns have place in the list; (Lakewood, Ohio; Plymouth, Pa.).

58. The college and high school groups are considered apart, since the former is of Massachusetts, the latter of Ohio origin. This affects the results as shown in the following tables:

Town.	Muskegon	Ansonia	Walla Walla	Hagerstown	Leominster	Cambridge	Aurora	New Castle	Duquesne	Newton	Huntington	Boisé	Washington (D.C.)	Washington (Pa.)	Erie	Berkeley	Topeka	Alexandria
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Correct, % College.....	11	5	13	3	73	97	12	6	4	93	0	17	94	3	37	57	59	4
High School..	15	0	8	0	2	81	6	17	13	0	74	21	81	8	60	9	49	0
Incorrect, % College.....	15	11	15	8	3	3	25	60	44	1	27	19	28		53	24	17	11
High School	19	15	24	19	7	9	91	76	21	42	13	34	30		15	30	15	32
Omitted, % College.....	74	83	72	89	23	0	63	34	38	5	73	64	23		9	19	24	85
High School.	36	85	68	81	91	9	4	8	66	58	13	45	37		25	60	36	68

Town.	South Bend	Paterson	Jackson (Mich.)	Jackson (Miss.)	Jackson (Tenn.)	San Francisco	Norwalk	Michigan City	Camden	Rock Island	Chattanooga	Malden	Pottsville	Belleville	Oswego	Kenosha	
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	AV.
Correct, % College.....	29	69	4	16	11	97	13	1	40	9	53	96	11	3	33	4	32
High School.....	64	36	8	26	11	98	0	0	26	17	85	2	8	2	13	0	25
Incorrect, % College.....	21	9	..	42	..	0	20	77	41	34	32	1	20	12	24	9	24
High School...	23	15	..	37	..	0	21	91	26	38	4	15	11	9	23	19	28
Omitted, % College.....	49	21	..	48	..	3	67	21	19	56	15	2	69	85	43	86	44
High School....	13	49	..	47	..	2	79	9	47	45	11	83	81	89	64	81	47

59. Provincialism of response, favoring now the college, now the high school group, is shown in the items Muskegon, Newton, Huntington, Erie, South Bend, Rock Island, Chattanooga and Malden. This is the most important special factor in the results. Separation of college from high school as such scarcely appears outside the items Berkeley, Paterson, Oswego. The Massachusetts group is unduly favored by random selection from the large list, through the number of Boston suburbs thus included. It is indicated that the test items should not include the States where the test is made, or perhaps the States adjacent thereto. Towns listed in more than one State are also inconvenient.

60. It is suggested that the ratio of right to wrong responses has an individual significance not included in the purpose of the test, concerning the reliability of the subject's "knowledge," his critical insight into the correctness of his answers. The measure used is the per cent which the number of right responses is of the whole number of responses written. Individual differences are as shown in the accompanying distribution:

Right Answers divided by All Answers	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
College Group (75)	1	1	2	5	4	8	12	9	6	6	5	3	6	2	2	1
High School Group (53)	2	0	7	2	7	6	8	8	4	5	4	..	..	..	..	..

61. Group difference is in the expected direction but is small in proportion to individual variability. The further education of the college group affects the criticism of this knowledge less than it does the capacity for cued combination responses.

62. Among the mechanisms of wrong responses, is first the existence of other towns having the same name as that properly designated. Jackson, N. H. (9)[6], and Camden, Maine (19) are prominent examples of this. The response Alexandria, La. (2) is virtually correct, 1915 census figures giving this town population of 18,000. (It is the only case of this kind.) Districts outside the United States are occasionally named; such responses may be fairly reasonable, as Alexandria, Egypt (2), or less so, like Topeka, China (2). Many varieties of autistic thinking are evident in the wrong responses, sound association among them. Duquesne is assisted into Iowa (7) by Des Moines and Dubuque. Boisé owes the same location (14) to the resemblance of Ia. and Ida. Jackson finds itself in Florida (59) beside Jacksonville. Norfolk brings Norwalk into Virginia (12). Rock Island proceeds via Rockland, into Maine (20), and via Block Island to New York (12); it is also found in Rhode Island (3). A peculiar feature in the results is the placement of Chattanooga in New York (16), apparently through such media as Chautauqua, Canandaigua, Cattaraugus, Canastota. Other than directly sensory associations appear responsible for the placing of Washington in Oregon (14). Berkeley, Mass. (16), has a partial determinant in the proximity to most of the subjects of Berkeley Street, Boston. Michigan City appears regularly in Michigan (112); but once out of 115 responses by 140 persons, in Indiana.

63. There is no reason to regard such material as an "intelligence" test. It applies synthetically to a restricted field of practical information. Illogical, "autistic" thought-processes are so prominent in the wrong responses as to suggest, analytically, that these reflect a capacity for critical recognition of the "real."

C. M. KELLEY,
F. L. WELLS.

[6] i. e. the response Jackson, N. H., was given nine times.

VI. AN OMNIBUS TEST FOR REASONING AND IMAGERY.

64. Performance in the present test-series, (R) is limited by reasoning and control of imagery rather than by range of information or attention to mechanical detail. The form used is reproduced herewith. The first of its tests has had con-

College Group				High School Group			
Point Score	Per cent Surpassed by same	Time of Test	Per cent Surpassed by same	Point Score	Per cent Surpassed by same	Points Earned per min.	Per cent Surpassed by same
21	0	27	0	0	0	0	0
30	0	21	1	20	6	2.5	17
40	1	19	1	30	17	5.0	43
50	9	17	7	40	27	7.5	73
60	23	15	9	50	38	10.0	91
70	24	13	22	60	53	12.5	96
80	45	11	46	70	57	15.0	98
90	66	9	74	80	72		
100	95	7	91	90	85		
		5	98	100	98		

siderable development elsewhere. The latter two are derived from the Stanford scale. It was desired to examine their adaptability to group testing. Ten alternative forms of the series were prepared.

65. In the clock tests the subject should not have access to a watch or clock face. One group of the high school students was able to see a large clock during the experiment. Their results are not improved over the others.

66. Scoring and recording were as described in paragraphs 41, 42, 51. Percentile scores for the college and high school students are as follows:

67. There is less difference of the college and high school groups than in series K, and more than in series C. As in series K, separation is more marked in total points earned than in points earned per minute.

1. *In each line every number differs from the one before it by the same amount. On each dotted line write the number that will come next in the same way.*

280, 273, 266, 259.....

378, 386, 394, 402.....

276, 280, 284, 288.....

232, 224, 216, 208.....

2. *Figure in your mind where the hands of the clock are when it is* 3:36 *o'clock. Now when the hands have changed places (so that the small hand has gone where the large hand was, and the large hand has gone where the small hand was), what time is it then?.....*

If it is	*What time is it when the hands change places*
8.13	.....
12.41	.....
10.19	.....

3. *A doctor must measure out exactly 3 ounces of medicine for a man who is sick in the woods. He has only a 2 ounce measure and a 5 ounce measure. Show how he can use these measures to get just the right dose of 3 ounces without any guessing. He pours from one measure to the other, and what he does not want, he pours back into his medicine bottle.....*	*How would he measure out 2 ounces with a 3 ounce measure and a 4 ounce measure.*
.....	.....
.....	.....
.....	*How would he measure out 1 ounce with a 2 ounce measure and a 5 ounce measure.*
.....	.....
.....	.....
.....	.....

68. Accompanying table analyses the credits earned in all data obtained with this series:

Performance	Test 1	2	3
Per cent Making Perfect Score	61	50	25
½ error	..	..	22
1 "	17	32	15
1½ "	..	..	3
2 "	7	7	5
3 "	3	4	26
4 errors, and omissions	12	7	4

69. No group difficulty appears with the clock test beyond preventing the consulting of watches and clocks. In test 3, (ingenuity), failures are largely due to "guessing", i. e. such responses as "fill the 3 oz. bottle half full," and the like. It may be questioned whether more emphatic instructions against this should be given. It is believed that to be satisfied with a "guessing" response in the present experimental setting, denotes essentially significant failure to grasp the problem. The test involves more writing than is desirable, otherwise is suitable for group presentation. It shows some notable separation of college and high school groups, perfect score being made by 38 per cent of the college and 4 per cent of the high school group.

C. M. KELLEY,
F. L. WELLS.

VII. CERTAIN INDIVIDUAL RECORDS COMPARED WITH STANDARD FILE.

70. In the previous studies the chief concern has been with college and high school students, that is with persons who have not in general made an independent adjustment to life. In the present study a comparison of their records is made with a few individuals who have made such adjustments, at various levels. The findings throw light on what capacities in these test functions are compatible with different levels of vital accomplishment.

71. Attention is invited to records of the following six subjects:

		Series C		Series K	
	Subject	Total points	Points per minute	Total points	Points per minute
Per cent surpassed by...	I	6	11	10	63
" ...	II	9	28	59	85
" ...	III	18	54	97	99
" ...	IV	56	87	41	75
" ...	V	85	96	52	96
" ...	VI	93	..	100	..

72. Subjects I-III are experienced clerks working in the same office. To their employer they rank in business performance in the order represented by their test results, but these do not directly express the differences between them. This is partly because the comparative material is generally selected from the same level, of intelligence in which these subjects fall. This "dilates" the percentile differences. If the standard file had contained material from all levels of intelligence instead of from a rather high level only, these percentile scores would be higher and closer together.

Subject IV is an ambitious young man of foreign birth, an office assistant, who has some language difficulty. Subject V is a college and medical graduate; Subject VI a Phi Beta Kappa college graduate, of further academic experience.

73. These cases make relatively higher scores in points per minute than in total points earned. This betokens more hurried work. The examinations in these cases were individual, tending probably to greater haste through greater self-consciousness. It is to be supposed that the attitude of the students was influenced by the analogous situation in academic examinations and quizzes, which are less acutely hurried than psychological tests. Money prizes were offered

for the best performance in the college and high school groups, but not in the other cases.

74. Five of the more successful type of insurance salesmen, together with an office manager, were given the tests by their chief executive. The percentile scores are as follows:

		Series C		Series K		Series R	
	Subject	Total points	Points per minute	Total points	Points per minute	Total points	Points per minute
Per cent surpassed by	VII	50	12	98	41	38	32
"	VIII	5	4	46	53	5	3
"	IX	27	7	28	12	68	82
"	X	31	26	44	11	8	11
"	XI	59	55	83	52	100	96
"	XII	75	94	86	83	84	80

75. Comment of the executive making the tests, based upon the points per minute score, is in part as follows:

"If we eliminate XII, who is an office man, we find XI standing high in each of the different tests. This is in accord with what I believe to be the relation of XI's ability to the other men. In fact, I should grade them about as follows, from the standpoint of salesmanship and general usefulness to the Company in that particular: XI, VII, VIII, IX.[7]

"The difference between the five men would not be very great, although XI, perhaps, stands out somewhat more conspicuously from the group than any of the others . . .

"On the whole, it would appear that the R tests come nearer to matching up with our judgment of the men than either of the other two. Even here, however, IX takes second position, but that, after all, may be nearer right than my judgment. Five years ago, I would have placed IX at the head of the entire group, but during the last five years the character of his work has changed very considerably.

"But even in the R tests, the differences seem to me to be very much greater than actually exist. It could not be said, for example, that XI is anything like so many times as good a man as VIII[8] when it comes to the question of his value to the Company.

"The C tests work out quite well as a measure of carefulness. XII, the office man, stands at the head of the list, while the other men grade downward exactly in accord with what I believe to be the fact . . . while the R tests, if

[7] A previous rating by the same individual is XI, VII, VIII, X, IX. A rating based on 1918 new insurance produced is VII, X, XI, VIII; IX's position cannot be clearly stated, but seems much improved. This has comparatively little correspondence with the test findings, (1917).

[8] Cf. par. 72.

looked upon as a measure of resourcefulness, seem to be quite accurate with the exception of VIII, who in that quality should, I believe, rank ahead of X and perhaps ahead of VII."

76. Only partial explanation is offered of the large differences in performance between relatively equal levels of nonexperimental competence. *Responses in the tests depend on what* is at the surface of the conscious, or can be brought there with such effort as the subject will apply thereto. Intelligence is more closely related to the content of the conscious than to emotional or motor processes; yet there are possible many differences in the content of the conscious among persons of similar degrees of intelligence. Military psychology has shown this on a large scale, for the content of the conscious as influenced by occupation. Occupational differences can be but little invoked in the present instances. Other determinants, affecting consciousness and personality independently of intelligence, are to be sought in temperament, general interests, attitude toward the tests. These appeared especially important with Subject I who is much nearer II and III in professional usefulness than the test performance indicates.

77. Intelligence is a factor in most civilized operations, and is quicker and surer reached in formal tests than otherwise. "Other things being equal" the more intelligent will always do better mental work; and as other things are equal in chance, formal tests are at all levels better than no tests at all. They are most effective where the desideratum is a more or less routine efficiency of definite mental associations, as in clerical work. They are less effective for operations that deal with things rather than with ideas, and still less effective at the more executive levels of endeavor where less depends on intelligence in proportion to such qualities as coöperativeness, persistence, initiative, tact. "Other things being equal" they will still pick the more useful individuals; but other things are more likely to be unequal, and their inequalities more important, than at routine levels of mental work.

F. L. WELLS.

NOTES AND NEWS

SHERWIN CODY. *Commercial Tests and How to Use Them.* World Book Company, Yonkers-on-Hudson, N. Y., 1919, p. 216.

This book presents the history and technique of the National Ability Tests, which were used as the basis for the Efficiency Employment Register of high school graduates now being offered by the United States Employment Service in New York City. These Tests are intended for prospective office boys, general clerks, sales people, stenographers and bookkeepers. They are arranged in two parallel series of equal difficulty, each series containing 19 tests on such topics as: tabulation, invoicing, fundamental arithmetic, business arithmetic, writing and answering letters, stenography, copying on typewriter and on mimeograph, addressing and filing envelopes, memory, English grammar, punctuation and spelling. Minute directions for administering the tests, scoring and computing the data and sample records from several school systems are given. The book should prove very helpful to those interested in educational and vocational guidance.

ARTHUR S. OTIS. *Manual of Directions for Administering the Otis Group Intelligence Scale, A Point Scale,* Edition 1919, Forms A and B. World Book Company, Yonkers-on-Hudson, N. Y., 1919, p. 37.

This little pamphlet gives detailed directions for administering the Otis group tests, the underlying principle of which was incorporated in the United States Army tests. The complete material for giving the tests consists of (1) Examination Booklets, Form A and Form B; (2) Individual Record Card, for recording all available information about the subject tested; (3) Log Slip; (4) Examiner's Manual; and (5) Examiner's Key. The Booklets contain ten tests each of parallel difficulty, each test being suitable to subjects from the fourth grade up through the university. The tests include the following items: (1) following directions, (2) opposites, (3) disarranged sentences, (4) proverbs, (5) arithmetic, (6) geometric figures, (7) analogies, (8) similarities, (9) narrative completion, and (10) memory. Each test is arranged so that its successive elements become more and more difficult, and the score is based simply upon the total number of elements correctly answered.

WILLIAM D. LEWIS and EDGAR A. SINGER. *The Winston Simplified Dictionary, Including all the Words in Common Use, Defined so* that they can be easily Understood. (Illustrated.) The John C. Winston Company, Philadelphia and Chicago, 1919, pp. xxii+820.

The vocabulary included in this book has been chosen with special reference to educational needs and the definitions are given in plain, simple language, avoiding the frequent dictionary error of repeating in the definition the word to be defined. One or two random samples will suffice to indicate the character of the work:

Calculate (kăl'kūlāt), *v. t.* to add, subtract, multiply, or divide any sum to find the result; to determine by any process of reasoning; to estimate, as, no one can *calculate* the benefits of electricity to the world: *v. i.* to make a computation; to rely, with *or* or *upon.*

Salm-on (săm'ŭn), *n.* a sea fish, found in northern waters, which ascends rivers to lay its eggs; the yellowish-pink color of salmon flesh; *adj.* of the yellowish-pink color of salmon flesh.

The Introductory Part contains an eight-page discussion of the topic "How Our Language Has Grown," ten rules for spelling derivative words, explanatory notes on varied spelling, capitals, hyphens, etc., a short list of abbreviations used, and a guide to pronunciation. It is greatly to be hoped that this new book will supplant the cumbersome school dictionaries now in use, at least in the elementary grades, and we recommend it not only for every classroom, but for every home as well.

The General Education Board has just published a series of eight reports on various aspects of the Gary Public Schools, including the following:

ABRAHAM FLEXNER and FRANK P. BACHMAN. *The Gary Schools: A General Account*, p. 265.

GEORGE D. STRAYER and FRANK P. BACHMAN. *Organization and Administration*, p. 126, with 7 School Programs.

FRANK P. BACHMAN and RALPH BOWMAN. *Costs, School Year 1915-1916*, p. 86, with 9 Tables and 22 Schedules for Expenditures.

CHARLES R. RICHARDS, *Industrial Work*, p. 204.

EVA W. WHITE. *Household Arts*, p. 49.

LEE F. HANMER. *Physical Training and Play*, p. 35.

OTIS W. CALDWELL. *Science Teaching*, p. 125.

STUART A. COURTIS. *Measurement of Classroom Products.*

The Colorado State Teachers College Bulletin of October, 1918, consists of "A Study in Spelling," by Professor J. D. Heilman. The Bulletin of March, 1919, contains "A Study of the Mechanics of Reading," by the same author.

The Committee on Classification of Personnel in the Army has issued "A Summary of the Work of the Classification Division of the Adjutant General's Department" which contains brief descriptions of the following items: The Task of Army Personnel Organization, The Classification of Soldiers, Soldiers' Classification Card, Trade Specifications and Occupational Index, Trade Tests, Tables of Occupational Needs, Personnel Specifications, Classification and Rating of Officers, The Rating Scale, The Rating Scale Card, Relations with the Field, and The Recruit Receiving Building.

In a recent *Bulletin of the University of Wyoming* Professor June E. Downey contributes a study entitled "The Will-Profile, A Tentative Scale for Measurement of the Volitional Pattern." The profile is based upon a quantitative measurement of the following ten factors: coördination of impulses, accuracy, tenacity, resistance, assurance, motor-impulsion, speed of decision, flexibility, freedom from inertia, and speed of movement. Most of the items are determined by an examination of handwriting under various conditions; as: normal, speeded, retarded, disguised, imitated, blocked, with eyes closed, with distractions by counting by 2's or by 3's, etc. Each of the ten tests is scored on a basis of ten points, and the will-profile for each individual is constructed on the scores obtained in the ten factors.

The Bureau of Educational Experiments in New York City has secured the services of Dr. William R. P. Emerson, under whose direction a series of experiments with "nutrition classes" have been carried out. Some of the results and conclusions of this work are presented by David Mitchell in an article on "Malnutrition and Health Education" in the recent March issue of the *Pedagogical Seminary*. This paper contains a number of valuable tables on the distribution of children according to percentage of over and underweight in grades I, V, VI, VII, in a special class for exceptionally bright children, and in an open air class. After an initial measurement in height and weight these classes were subjected to several kinds of treatment; for instance, the 7th grade received instruction in health habits, physical care in the removal of defects interfering with nutrition, a half-hour rest period during the forenoon, and a mid-day lunch. The other classes received similar combinations. At the end of 19 weeks the children were remeasured to observe the changes produced by the various methods of treatment. Among the causes for lack of gain or loss in weight the following were prominent: enlarged tonsils and adenoids and carious teeth. Some of the important conclusions drawn by the author are: that provision of food in school is in itself not adequate to solve the problem of malnutrition, that stimulants, such as tea and coffee, should be eliminated, that physical defects should be removed, and that the children should be taught in hygiene and health habits.

The *Thirteenth Biennial Report* of the Board of Trustees of the Preston School of Industry, Ione, California, covering the time from July 1st, 1916 to June 30, 1918, contains an enquiry into the mental conditions, character and antecedents of 300 boys who were in the school on June 1st, 1918. The psychological data were collected by Fred. H. Allen and Warner Brown and were based upon a mental examination with the Goddard revision of the Binet scale and other supplementary tests, and upon a critical behavioristic study made during the first few weeks of the boy's stay at the school. On the basis of this study it was found that about one third of the 300 boys were dull normal or above, one third borderline cases, and the remaining third were definitely defective. Of the latter group, 27 cases showed symptoms of slight psychopathic instability and 6 cases presented conspicuous symptoms of mental abnormality.

The National Association of Employment Managers began last January the publication of a monthly bulletin, called "*Personnel.*" The second issue contains a succinct description of trade tests. Three columns in each issue are devoted to recent publications in related fields. We welcome this new publication and trust that it will contribute materially to a widespread recognition of the human factor and the psychological problems involved in the relations between employer and employee.

The University of Illinois has established a Bureau of Educational Research under the directorship of Professor B. R. Buckingham. Among the aims of this Bureau are: the compilation of bibliographies on educational topics; assisting school authorities in devising records for general and specific purposes; the study of such problems as the grouping of grades, the function of special classes, continuation and part-time schools, school statistics, and school accounting; giving aid

in selecting, retaining, supervising the teaching force; and the administration and computation of educational tests in coöperation with local school authorities.

The scope of the Bureau of Salesmanship Research at the Carnegie Institute of Technology, Pittsburgh, has recently been enlarged to include factory and clerical workers and executives, as well as salesmen. The organization is now known as the Bureau of Personnel. Its purpose is to study methods of selection, training, development, supervision, and compensation of employees.

The Public School System of Santa Ana, California, has established a Department of Research, with Mary Bess Henry as Director. The first Bulletin published by this new Department contains a discussion of "Mental Testing as an Aid in Guidance and Classification of School Children." The results reported are based on Otis' vocabulary tests, Otis' group tests of intelligence, teachers' estimates of intelligence, and scholastic records.

Professor L. L. Thurstone of the Division of Personnel and Psychology at Carnegie Institute of Technology, Pittsburgh, has recently issued two booklets for the psychological examination of college freshmen and high school seniors, Part A and Part B, and also a "Test of Engineering Aptitude for College Freshmen and High School Seniors." The former consists of one hundred short problems with directions for each. The subject is asked to solve as many as he can in twenty minutes. The engineering test contains 24 mechanical problems with directions, the subject having 30 minutes for solving as many as he can do.

Dr. Walter B. Swift, has just returned to Boston from Cleveland, where he has spent about a year installing and supervising Speech Correction in the Cleveland Public Schools. He trained up 15 teachers who are part time speech teachers. They do their regular grade work as usual. The part time speech teacher is one of the unique features of the Swift methods and systems of speech correction. In Cleveland there are now 46 classes and over 600 cases under treatment. Dr. Swift will be in Milwaukee, Wis., in July and Athens, Ohio in August. He is booked to give informal courses in these two cities this summer.

The National Society for the Study and Correction of Speech Disorder will have its summer meeting in Milwaukee, on July 4, as one of the affiliated Societies of the National Educational Association. Members of the Society and invited guests of prominence in the field of speech correction, will address the Association. Anyone interested to receive an advanced program may do so by addressing the Secretary, Miss Marguerite Franklin, 110 Bay State Rd., Boston Mass.

The following publications have been received:[1]

James Burt Miner. *Deficiency and Delinquency, An Interpretation of Mental Testing.* Educational Psychology Monographs, No. 21. Warwick and York, Baltimore, 1918, p. 355.

Roy Franklin Richardson. *The Psychology and Pedagogy of Anger.* Educational Psychology Monographs, No. 19. Warwick and York, Baltimore, 1918, p. 100.

---

[1] Mention here does not preclude further comment.

The Journal of Educational Psychology, Vol. x, No. 2, February 1919. Published by Warwick and York, Baltimore.

University of Oregon Bulletin, New Series, Vol. xv, No. 12. The University High School, issued by Department of Education. September 1918. Published by University Press, Eugene, Oregon.

Children's Year Leaflet No. 10. Advising Children in their Choice of Occupation and Supervising the Working Child. U. S. Department of Labor, Children's Bureau. Bureau Publication No. 53. Published by Government Printing Office, Washington, 1919.

Carry On, A Magazine on the Reconstruction of Disabled Soldiers and Sailors. Edited by the Office of the Surgeon General, U. S. Army. Published by the American Red Cross. Vol. 1, Nos. 7 and 8, April 1919 and May 1919.

The Porto Rico School Review. San Juan, Porto Rico. Vol. 3, Nos. 6 and 9, February and May 1919.

Revista de Psiquiatria y Disciplinas Conexas. Vol. 1, No. 3, January 1919. Published at Gremios 435, Altos-Lima, Peru.

Revista de Educacion. Tercera Epoca. Vol. 1, No. 2, March 30, 1919. Published by the Servicio Nacional de Instruccion Publica. Santo Domingo, Republica Dominicana.

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. III SEPTEMBER, 1919 No. 3

A NON-LANGUAGE GROUP INTELLIGENCE TEST

By RUDOLF PINTNER, Ohio State University

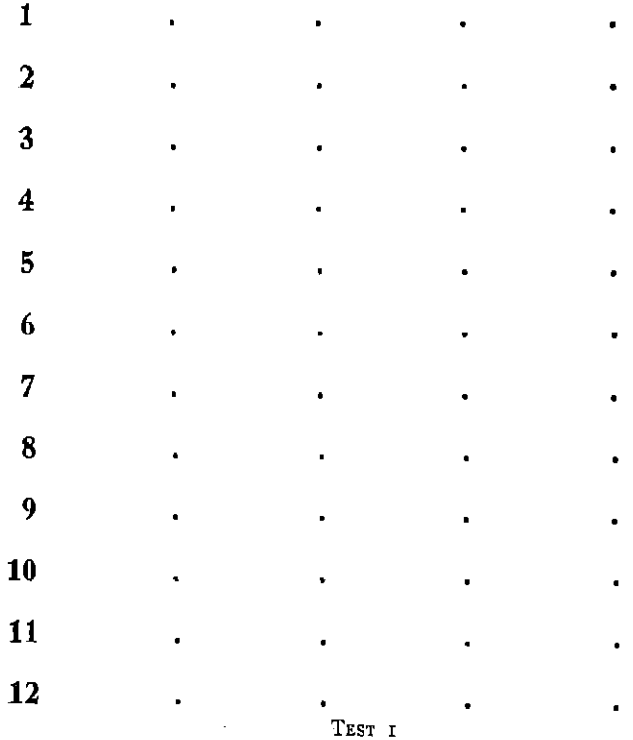
The usefulness of group tests for the measurement of intelligence is resulting in the appearance of a great number of such tests, and the development of this type of test marks the most important advance in mental testing during the past few years. The best known group tests at the present time depend largely, if not entirely, upon the knowledge and use of language. They presuppose the ability to read and write the English language. Apart from the debatable question as to whether ability to use and handle words is the best measure of general intelligence or as to whether it alone should be used as such a measure, it is very desirable to have tests which do not involve any language. In our industries and schools there are many foreigners whose knowledge of English is entirely lacking or so meagre as to be useless for testing purposes; and in addition we have a great many illiterates. The Army Beta Test and the recently published Thorndike[1] Group Test without Language are the only tests adapted for such subjects known to the writer at the present time.

For the past two years the writer has been working with a set of tests that involve no language, that can be given to illiterates, that do not demand a knowledge of English in order to understand the directions for doing the tests. In this way the illiterate, the foreigner and the deaf will all be given an equal chance with the hearing English-speaking literate individual.

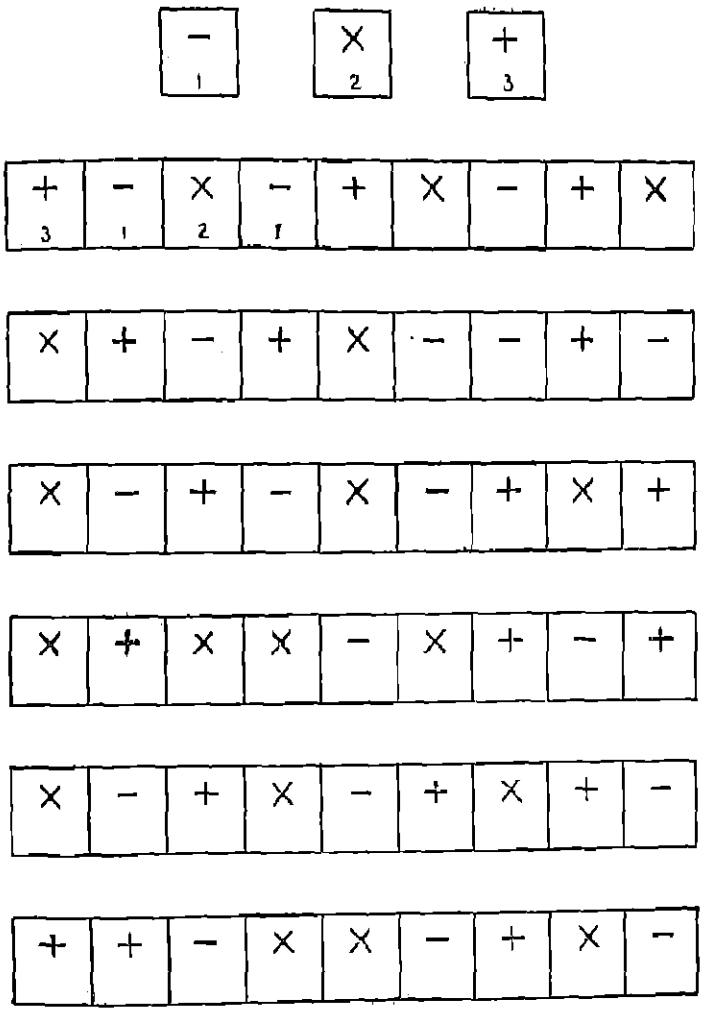
[1] Thorndike, E. L. A Standardized Group Examination of Intelligence Independent of Language. *Journal of Applied Psychology*, Vol. III, No. 1, March 1919. pp. 13-32.

Test No. 1.—The Imitation Test.

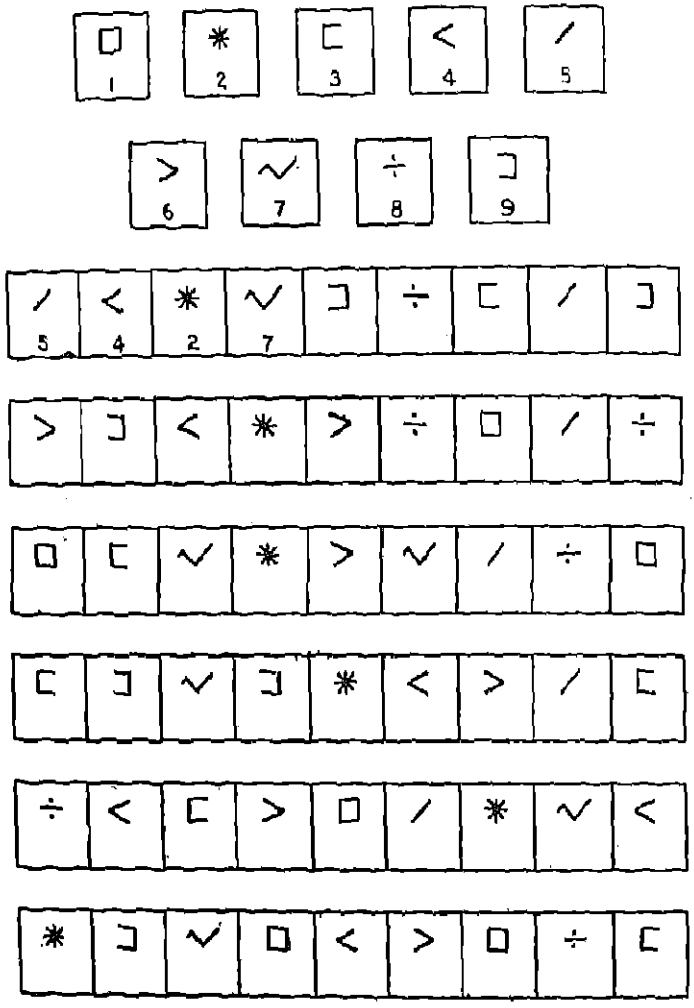
This is essentially the same as the Knox Cube Test using dots on a blackboard and a pointer to make the moves. The subjects draw lines from one dot to another showing the sequence and direction of the moves. The twelve lines in the test series were chosen out of a trial series of twenty-five and they progress gradually from easy to hard. The twenty-five lines of the trial series were chosen in such a manner that the drawing of the lines from one dot to another would be easily distinguishable on paper. In some combinations of movements from one dot to another lines coincide, and it would be impossible, when correcting the paper, to tell whether the subject had been working correctly or not. The trial series of 25 lines was given to 245 university students and also to 56 school children in the fourth and sixth grades. The following is the trial series of twenty-five lines, together with the percentages of correct responses for the children and the adults.



TEST 1



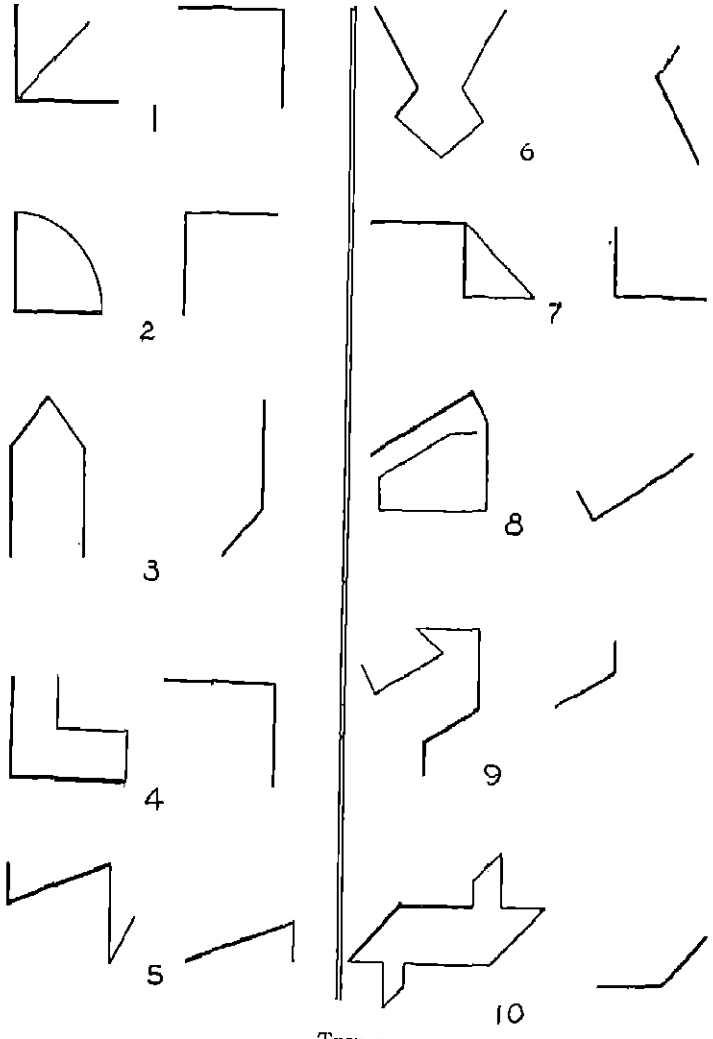
TEST 2



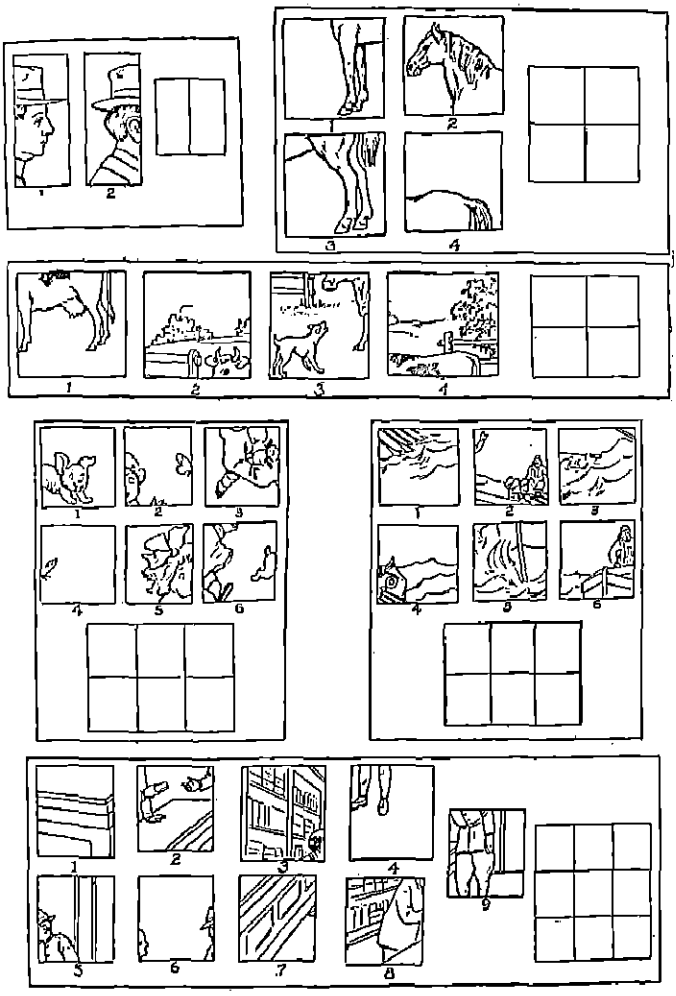
TEST 3



TEST 4



TEST 5



TEST 6

No.	Line	Per cent Correct. Adults	Children
1	1 2 3 4 2	89	86
2	1 2 3 4 1	98	97
3	1 3 4 2	88	64
4	1 2 4 3	79.5	57
5	1 3 2 4	88	61
6	1 4 2 3	93.5	52
7	1 4 3 2	96	70
8	1 2 4 3 1	65	29
9	1 3 4 2 4	57	27
10	1 4 2 3 4	81	41
11	1 3 2 4 1	74	38
12	1 2 4 3 2	68	34
13	1 3 4 2 1	71.5	41
14	1 4 3 2 1	94	66
15	1 3 2 4 2	71	23
16	1 4 2 3 1	80	20
17	1 3 2 4 3 1	62	13
18	1 2 4 3 1 4	45	14
19	1 3 4 2 1 2	54	14
20	1 4 2 3 1 2	50	7
21	1 4 3 2 1 2	78	23
22	1 2 3 4 2 1	80	50
23	1 2 3 4 3 1	90	46.5
24	1 2 4 3 2 1	77	43
25	1 2 3 4 1 4	91	59

The correlation between the rank order of the lines for the adults and children is .87. This correlation is very high and, therefore, it is somewhat immaterial which order we take. The greatest discrepancies in the difficulty of the lines occur in line 16 where we have a difference of 9.5 rank places, in line 21 where we have a difference of 5.5 places and in lines 6, 23, 1 and 4 where we have a difference of 5 places.

From these twenty-five lines we have chosen twelve which gradually increase in difficulty, basing this largely upon the adult order of difficulty but taking into account the order of difficulty obtained from the children's records. The final series of lines chosen is given under the directions for giving the tests.

Test No. 2—The Easy Learning Test.

This test is simply a very easy substitution test of the type common in psychological tests. The idea was to make a very simple substitution test which very young children could manage. For this reason only three digits were used and three very easy symbols. The squares are large enough so that small children can write in digits, and the sample digits filled in at the beginning of the test give the child an idea of what he is supposed to do.

Test No. 3.—The Hard Learning Test.

This is essentially the same as the previous test and in fact the Easy Learning Test serves as a method of teaching what is required in the Hard Learning Test. In the Hard Learning Test the child performs the same operation as in the easy test, the only difference being that here he has to work with nine digits and symbols instead of with three. The digits and symbols are those used in the standard Symbol-Digit Test.

Test No. 4.—The Drawing Completion Test.

This is a shortened form of the large Drawing Completion Test devised by Pintner and standardized by Pintner and Toops.[2] The series of pictures is arranged in order of difficulty from very easy to relatively hard.

Test No. 5.—The Reversed Drawing Test.

This test was suggested by Mr. H. A. Toops. We made a preliminary experiment with a large number of drawings and tried these out on a great many observers before choosing the ten for the final series. The final series was again arrived at by choosing drawings that progress in difficulty from easy to hard. In each case the figure to be completed is started by two lines so that whether the reversal is up-down or right-left the observer must himself decide by finishing the picture from the two lines that are given. It is obviously a type of spatial relations test.

Test No. 6.—The Picture Reconstruction Test.

This test was devised by the writer, the idea being to give as a group test the common operation of putting blocks together that will form a complete picture. It is essentially the same as the child's game of putting pieces together to form a picture. However, when we transfer this to paper and do not allow the manipulation of the parts of the picture, it becomes very much more difficult. It means that we have to reconstruct more or less in our minds the complete picture. Several pictures were arranged and mimeographed copies of these were made from which six were chosen. These six progress from very simple to very hard pictures.

The whole series of six tests are made up in booklet form allowing one page for each test and the first cover page for name, age, school, etc.

[2] Pintner, R. and Toops, H. A. A Drawing Completion Test. *Journal of Applied Psychology,* Vol. II, No. 2, June 1918. pp.164-173.

Time Limits.

After the final series for each test was decided upon, it was necessary to fix upon a time limit for each test. For this purpose the tests were tried out on adults and children in order to get some idea of adequate time limits. As a general rule a time limit for any test should be such that none, or at least very few, should be able to finish within the time limit. Since these tests are to be given to all ages from six or seven upwards it is, of course, impossible to adhere to this standard rigidly, because, if we did so, we should have to fix the time limit in accordance with what the quickest adult could perform. In many cases this would mean that the seven-year-old child would hardly have time to begin the test before the time limit would have elapsed. We must, therefore, adjust our limit rather to the medium aged child.

Furthermore, we can draw a distinction between some tests where the time is very essential and others where rapidity of work is not so essential. Rapidity of work is the greatest factor in tests two and three, but is not nearly so important in the other tests.

A group of children were tested and the time for the first who finished was taken. This gives us the following times:

Test 2—47 sec.
Test 3—73 sec.
Test 4—1 min.
Test 5—2.5 mins.
Test 6—5 mins. (but not correct).

For Test 2 it took 1 minute 20 seconds for about half the class to finish, for Test 3, 1 minute 45 seconds for half the class, and Test 4, 1 minute 45 seconds. In Tests 5 and 6 the children were stopped after 6 minutes because practically none of them had finished and these are tests over which children may ponder for a long time.

Using this data and other results from adult observers as a basis the time limits, as given below under the directions for giving the tests, were decided upon, making a total of 13.5 minutes in addition to the time required for giving Test 1.

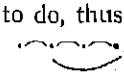
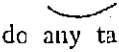
With reference to these time limits we may note that one minute for Test 2 is rather too long for older children, but since this test is an important one for younger children it is necessary to give them adequate time to get under way. It takes an adult working rapidly about 37 seconds to finish, but only 50 to 60% of University students finish within one minute. For Test 3, 1½ minutes will permit very few chil-

dren to finish. In Test 4, 2 minutes as a rule is ample, and this length of time is allowed because rapidity of work is not the essential factor in this test. The time limits of 4 and 5 minutes for Tests 5 and 6 respectively are very liberal for the older children and this has been purposely so arranged, because in these two tests quality of work rather than speed is the essential factor. With the younger children these time limits have been found to be fully adequate for all that they are able to accomplish.

Directions for Giving the Tests.

Pass papers—have subjects put full name, age last birthday and grade on first page and tell them not to turn page until told to do so.

Test 1.—Imitation.

No time limit. Draw on blackboard one row of four large dots and also to one side of the board three rows of smaller dots in rows as on sheet. Move the pointer from dot to dot in order 1 2 3 4 and, using the first series of smaller dots, draw lines to show subject what he is to do, thus  . Follow with samples 1 2 3 4 2  , 1 4 2 3 1  . There is no need to do any talking if the children are foreign or deaf. After giving the samples have the subjects take their pencils. Erase the three samples on the board.

Give the following series in order with sufficient intervals for the subjects to draw the connecting lines.

1.....	1 2 3 4 1	7.....	1 3 2 4 1
2.....	1 4 3 2	8.....	1 2 3 4 2 1
3.....	1 4 2 3	9.....	1 3 2 4 2
4.....	1 4 3 2 1	10.....	1 2 4 3 1
5.....	1 2 3 4 3 1	11.....	1 3 4 2 1 2
6.....	1 3 4 2	12.....	1 2 4 3 1 4

Before beginning each row be sure to have the attention of the group and at the end give some signal to them to draw the line on their paper. Move the pointer from one dot to the other without raising it from the blackboard at the rate of about one a second pausing somewhat on each dot.

Test 2.—Easy Learning. Time limit 1 minute.

Hold sheet up before the class and point to the key at the top, first at the symbol, then to the numbers, explaining that they must copy in the numbers. Also show the children that

the first few squares have been filled in and indicate that they are to complete the sheet.

With young children draw a sample on the board.

Test 3.—Hard Learning. Time limit 1½ minutes.

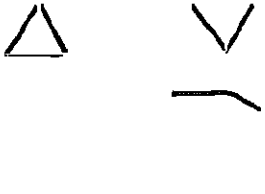
Proceed as in Test 2.

Test 4.—Drawing Completion. Time limit 2 minutes.

Draw a head without one eye on the board. Ask what is missing. Draw in the eye. Hold the test blank up before the class and pretend to draw the eye, and indicate they are also to complete the other drawings.

Test 5.—Reversed Drawings. Time limit 4 minutes.

Draw on the board the two following samples, accentuating the heavy and fine lines:



Trace with the finger the heavy lines on the first figure and then the similar lines on the second figure. Cover the first figure with the opened left hand palm towards the board, turn the hand slowly over, at the same time bringing it to cover the second figure, showing by the motion that the figure has been reversed. Repeat this movement once or twice. Now trace the fine line on the first figure with the finger and show where it should go on the second figure. Then draw in the fine line and repeat once again the turning-over movement of the hand. In the same way demonstrate the third and fourth figures composing the second sample. If the children are hearing English-speaking appropriate language should accompany the demonstration, without however adding to the instructions.

Test 6.—Picture Reconstruction. Time limit 5 minutes.

For this test a sample picture cut in half and mounted on cardboard with the numbers 2 and 1 on the two halves of the picture must be prepared. The writer uses a drawing of a little girl traced in India ink on two pieces of white cardboard each measuring about 6 x 17 inches. The number 2 must be at the bottom of the left half of the picture (facing the picture) and the number 1 on the right half. Place the two pieces in order 1 2 on the rail of the blackboard, i. e., the two

parts of the picture will not be in the right order. On the board draw a square divided into two parts for the numbers as on the test blank. Trace with finger the outline of the first part of the picture and indicate into which space it would go on the board and similarly with the second part of the picture. Then point to number 2 and write it in the first space on the board and then to number 1 and write it in the other. Pointing to the numbers and spaces may be repeated more than once. Having written in the numbers now hold the two parts of the picture over the space on the board so as to make a complete picture and indicate the complete picture. Then show that the numbers in the spaces on the board correspond *with the numbers at the bottom of the two parts of the picture* which is now being held just above them. This demonstration is generally sufficient to prepare the children for the test, beginning as it does with a very simple example almost the same as the sample used in the demonstration.

Scoring the Tests.

Celluloid stencils are used in scoring Tests 1, 2, 3, 5 and 6, thus simplifying the work and making it very rapid. A stencil does not help in scoring Test 4 (Drawing Completion). The following crude scores are given:

Test 1—One point for each line absolutely correct. Total, 12 points.

Test 2—One point for each space filled in with the correct digit. Total, 50 points.

Test 3—One point for each space filled in with the correct digit. Total, 50 points.

Test 4—Two points for the right thing in the right place; one point for the right thing but not rightly placed. Total, 20 points.

Test 5—One point for each reversed drawing correctly performed. Total, 10 points.

Test 6—One point for each number in the proper space. Total, 31 points.

Having scored the test these crude scores are recorded on a record card and weighted as follows:

Test	Maximum Crude Score	Weight	Maximum Weighted Score
1	12	9	108
2	50	2	100
3	50	2	100
4	20	5	100
5	10	10	100
6	31	3	93
Total			601

Correlations.

The following correlations have been obtained between each test and the total score on the whole series of tests:

	Test						
Age	n.	1.	2.	3.	4.	5.	6.
9.....	127	78	83	84	76	67	65
12.....	110	67	78	80	50	50	56
University Students.....	74	36	44	45	41	66	66

Tests 2 and 3 give the highest correlations for the children, while Tests 5 and 6 give the highest correlations for the adults.

Standards.

Tentative standard scores based upon results with ordinary grade children according to age are shown in Table I, which gives the number of cases tested at each age, the median and the 80 and 20 percentiles. It will be seen that the median

Table I. Distribution of Scores by Ages

Age	7	8	9	10	11	12	13	Univ. Students
No.....	72	97	143	195	175	127	111	72
80 Percentile	123	231	301	338	352	357	382	549
Median.....	76	130	190	272	285	290	309	501
20 Percentile	38	72	111	170	202	202	209	419

score rises steadily at each age, and the same is true of the 80 and 20 percentile scores.

Table II gives a percentage distribution of the cases arranged according to age and to score, showing a steady in-

Table II. Percentage Distribution According to Age and Score

	Age								
Score	7	8	9	10	11	12	13	14	Univ. Students
0— 99....	67	33	18	12	4	2	5	..	..
100—199....	28	37	34	13	15	16	14	16	..
200—299....	5	25	26	38	37	34	27	33	..
300—399....	..	5	21	32	36	40	42	38	15
400—499....	..	..	1	5	6	7	12	13	34
500—601....	..	..	..	2	1	..	..	..	51

crease in score from age to age and a fairly wide distribution of abilities at each age, such as we would expect among unselected school children.

Figure I shows a percentage distribution curve for ages 7, 9, 11, 13 and University Students. We note how the curves shift over from the left to the right with the advance in age.

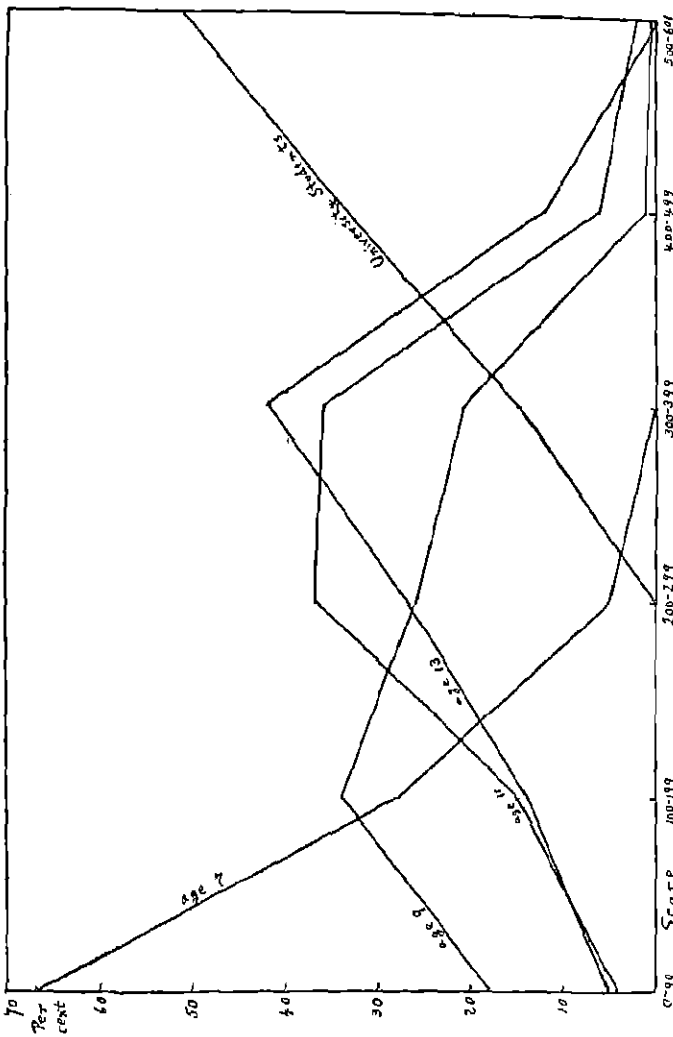


FIG. I. PERCENTAGE DISTRIBUTION. BY AGES.

None of the seven-year-olds make scores above 299 and none of the University students make scores below 300. Although the modes of the curves at each age are distinct and increase in score, there is considerable overlapping in abilities at each age as there should be.

The results given here can only be regarded as tentative standards. A much larger number of cases should be tested in order to arrive at accurate percentile norms.

So far only 16 cases have been tested on the Stanford Binet and this group test. The correlation between the I. Q. on the Stanford and the percentile rank on these tests is .66. There is obviously a fairly close agreement between the two ratings.

With regard to the scores of the University students it will be noted that about half of them score over 500 points out of a possible 601, but that there also is a fair scattering of cases down to scores of 300. None obtained the maximum score, so that it would seem that the test is capable of giving some kind of a differentiation among people of high intelligence. A group of nine university instructors were also given the test. The median score of this group is 573, as contrasted with the median of 501 for university students. Indeed all, except one, of the instructors made a score higher than the median of the students. It would seem, therefore, that the test might show some differentiation of ability among those making high scores.

Summary.

1. A set of six non-language tests has been arranged for group testing purposes, suitable for school children and for adults.

2. The tests can be given alike to English-speaking and non-English-speaking subjects, to literates and illiterates, to deaf and to hearing.

3. A method of procedure in giving the tests has been worked out. The time required for explaining and giving the tests is about thirty minutes.

4. Correlations between each test and the total score are all positive and fairly high.

5. Tentative norms have been arrived at for children of ages 7 to 13 inclusive, and for University students.

6. The tests should prove useful in mental survey work particularly in communities where there is a large foreign or illiterate element.

A GROUP INTELLIGENCE SCALE FOR PRIMARY GRADES

By Frances Lowell, University of Minnesota

Introduction

In October, 1917, the writer began a survey of all the schools in one of the representative counties in Minnesota, for the purpose of locating all cases of subnormality.[1] Such a survey necessitates, among other things, visiting every school in the county, selecting all children of questionable mentality, and testing them, individually, with the Binet-Simon tests.

The selection of retarded and subnormal children above the third grade is relatively easy, for teachers' estimates, previous school records, and repetition of grades without the excuse of a long continued absence taken together form a fairly accurate basis. However, such methods are practically impossible in and below the third grade. This is due to several factors: first, to the inaccuracy of the school records for primary grades; second, to the frequent change of teachers, especially in the rural schools; third, to the fact that many children have been in school too short a time to become sufficiently retarded pedagogically to suggest a diagnosis of mental deficiency; and fourth, to the inability of many children, on entering school, to do themselves justice in the grade because of the strangeness of the new environment. Other difficulties, common to all grades, are also encountered in such a survey. Teachers frequently persuade themselves that since a child is sweet, docile and attractive, he must necessarily be bright; or, because he is doing good work in his grade, he is undoubtedly normal, regardless of the fact that he is two or three years older than his classmates. In one rural school, the teacher, on being asked about the work of twin girls in the third grade, said with great assurance, "Oh, those girls are very bright. They do good third grade work, and they are such nice children." The twins in question had a chronological age of twelve years, a fact entirely overlooked by the teacher in her estimation of their ability. Mentally, they were found to be

[1] This survey was made under the direction of Dr. Fred Kuhlmann, Director of Research, Faribault, Minnesota.

7⅝ and 8⅛ years, respectively, when given the Binet tests. Hence, in order to have an accurate survey, the need for an impersonal method of selecting children of doubtful mentality was imperative.

History of Group Tests

At the time the survey was undertaken, no group scales of intelligence had been published. Many single group tests had been standardized and published, but these, alone, are insufficient for making even a rough classification of children. However, within the past year, two scales have appeared, namely, one given in the "Mental Survey"[2] and "A Group Point Scale."[3] The group tests used in army work are not discussed here because they are intended for adults only, and, too, because they have not been published.

Pintner, in his "Mental Survey," recognizes the need for standardized survey tests for selecting individuals for further examination and for special promotions in the schools. To provide such a scale, he uses well known tests, namely:

1) Rote memory
2) Digit-symbol
3) Symbol-digit
4) Word-building
5) Opposites
6) Cancellation.

Norms, in terms of percentiles for each chronological age from six to sixteen inclusive, are given for each of the six tests. Each individual's record of actual attainment is taken for each test, and later translated into its corresponding percentile. The index of a child's mentality is found by getting the median of the six percentiles.

The reliability of the survey scale is checked by correlating results with those found by using the Yerkes Point Scale. In the sixth grade a correlation of only .29 was found; in the fourth grade a correlation of .59 existed, and in the second grade, one of .71. When compared with results from Binet tests, the correlation was found to be .66. These results tend to show, according to Pintner, "that we must not place too much reliability on the survey tests for individual purposes, but that for groups they give a fairly accurate measure of mentality."

[2] R. Pintner: *The Mental Survey.* 1918.
[3] S. L. Pressey and L. W. Pressey: A Group Point Scale. *Journal of Applied Psychology,* September, 1918.

The Presseys took charge of the mental testing done in connection with the survey of a certain county in Indiana, which the Departments of Psychology and Sociology of Indiana University started in the fall of 1917. They, too, felt the need for group tests where large numbers of children were to be tested and so they tried to work out "a scale composed of tests which should be applicable over a wide range of ages —if possible from the second grade through high school— and which should be highly differential of general intelligence throughout this range."

The first requirement, in the selection of any test, is that its "beginning be easy enough and the directions clear enough so that nine-tenths of the 3B children would make some score, and that the end of the test should be hard enough so that no one of the high school sophomores (averaging sixteen years old) would make a perfect score."

The second requirement for the tests, and the one which the Presseys consider the most original feature of their scale, is that of "careful adaptation in the nature of the tests, and in flexibility and ease in control of the class, for work with school children."

The tests used are:

1) Rote memory
2) Logical selection
3) Arithmetic
4) Opposites
5) Logical memory
6) Word completion
7) Moral classification
8) Dissected sentences
9) Practical information
10) Analogies.

The results of these tests have been worked out in terms of percentiles, and norms published for about one thousand children. Then the reliability of the tests in sorting out extreme cases was checked up by comparing group records of twenty-one children found in the public schools—who, according to the Binet tests, had I. Q.'s of .76 or below—with records of normal children. It was found that the twenty-one cases "average at the 4 percentile of their age by the group tests; the highest scores made by these cases were at the 8 percentile."

The ability of the tests to sort out extreme cases was further noted in testing a group of "gifted" children with I. Q.'s above 1.25, and also a group of children from an institution

for feebleminded. The former group made scores averaging at the 98 percentile, whereas, in the latter group, only three were able to make a rating above the lowest 10 percentile of school children of the same ages.

Again, the age-grade status of children was compared with their rating by the group scale, and "the 143 retards in the ages from ten through sixteen average 70 points lower on their group test score than the 124 accelerates of the same ages."

The correlation of group test rank with teachers' rank, averages .65.

Finally, in order to show the reliability of the tests as a measure of general intelligence, comparisons have been made between results from three cities for one age and also for one grade. "The results obtained by the group tests by age show an average difference for the medians of less than two points, whereas the differences by grade average thirteen points." In other words, if a comparison of results from a single age group, as eight, is made for each of three cities, there will be an average difference for the medians of less than two points. However, if one compares the results from one grade, such as the third, for each of the three cities, the difference for the medians will average thirteen points.

Although both the Pressey and Pintner scales seem to furnish a fairly reliable method for sorting and classifying large groups of children above the second grade, neither of them can be applied in the kindergarten or first grade, because children just entering school can neither read nor write, and both of these accomplishments are necessary to success in the Pintner and Pressey scales. Therefore, even had these tests been available in the fall of 1917, they would not have been entirely satisfactory for the purposes of the writer.

Purposes of the Present Scale

In devising and standardizing the present group tests, the writer had two purposes in mind: first, to secure a scale for measuring the intelligence of large groups of children accurately enough to sort out all children of questionable normality. One could then be reasonably certain that these children, when tested individually by the Binet revisions, would be found mentally deficient, and the time ordinarily spent in testing cases wrongly estimated by teachers to be subnormal, would be saved.

The second purpose was to obtain a group scale which would discriminate between dull, average and super-normal

children in kindergarten and primary grades. One finds here children who, regardless of their chronological age, are mentally capable of attempting the school work of the next higher class. On the contrary, some children will succeed better if permitted to remain a second term in one grade, and receive special help in overcoming individual difficulties. Still others, though chronologically eligible, should be kept at home or in kindergarten, until their mental age justifies their entrance into the first grade.

The importance of correctly selecting and placing the children of various degrees of mentality at the very outset of their school life, is realized when one considers that a dull child who attempts the same work as a normal child becomes so discouraged by his constant failures, and so overwhelmed by the hopelessness of the task, that it takes years to overcome the resulting lack of self-confidence, and to make him exert his best effort; whereas, the gifted child, for whom the work of the first grade is too easy, forms unfortunate habits of study, becomes uninterested and often troublesome. Meanwhile, taxpayers bear an unnecessary burden when children have to repeat grades or are allowed to spend a whole year on work which might have been done in half the time.

So with the object in view of supplying a possible solution for the problems involved in school promotions, as well as the more immediate purpose of securing a group scale for sorting out subnormal children for the survey, the present series of group tests to be used in primary grades was devised.

Some of these group tests are adaptations of the original Binet-Simon tests; some have been taken from Dr. Kuhlmann's 1917 Revision and others from Dr. Terman's Revision; still others were suggested by tests standardized by the Bureau of Analysis and Investigation in New York; and the rest are original.

The Principles Involved in the Selection of Tests

The requirements of the group test are many, for not only must it possess the characteristics necessary to the individual test, but it must also satisfy various other demands. Simplicity is an important criterion in the selection of the group test: simplicity of material, of directions, of response, and of scoring.

Since a large number of children must be tested at one time, it is necessary that tests be selected in which the material used can be easily carried and quickly distributed. In order to facilitate handling, and to avoid confusion, the writer has ar-

ranged the material for the group tests in booklets. To each child, who has been in school less than one year, i. e., who is in the first grade, and hence can read and write but little, a booklet, 7½ in. x 7½ in., is given, which contains material and blanks necessary for the fifteen tests designated as five, six and seven year tests. Children in the second grade receive similar booklets designed for six, seven and eight year tests; while third grade pupils get booklets for the seven, eight and nine year tests. Thus, all the responses of each child are kept together, and confusion incurred by distributing and collecting loose sheets of paper every few minutes is avoided. All pages of the booklet are numbered, and definite spaces are allotted to each test. All possible elimination of disturbing elements is important, for in the primary grades the attention of the children is easily distracted.

Not only must the material for the tests be simple, but the directions to be followed must be clear and brief. Herein lies the greatest difficulty in selecting tests for a group. Frequently one feels confident that the directions for a certain test are perfectly clear and that they could not possibly be misunderstood, and yet when the test is given to a group of children, he finds that the meaning is entirely lost. Good English must frequently be sacrificed, for the child in the primary grades is surprisingly limited in vocabulary. Originally, in giving the directions for one of the six year tests, the writer said, "Make a cross in the LARGEST square." The result was puzzling, for the children crossed the smallest square as often as they did the largest. The test apparently was a failure for that age group. However, before discarding it, the writer decided to experiment a little to see if the difficulty could be discovered. Instead of having the children cross the square, they were asked to point to the largest square, whereupon one little girl tearfully informed the writer that she "didn't know what that meant." That solved the problem. From then on the children were instructed to "Make a cross in the BIGGEST square," and the success of the test was assured. Brevity is equally essential in giving directions to children, for their habits of thinking, of concentrating the attention, are not yet formed, and to listen, comprehendingly, to long directions is impossible for them. Novelty and change are so obviously important in maintaining the interest and stimulating the child to his best effort, that further discussion of the value of brevity is unnecessary.

In the selection of group tests, only those should be chosen which permit of but one correct response, and that response

must have but one possible interpretation. It is true that some of the tests consist of several trials of the same sort; thus, three different series of five digits each, are read to the children, to be reproduced, but the child passes the test if he succeeds in reproducing any one of the series correctly. This is a necessary procedure, because it is almost impossible, in a group, to have perfect silence while the directions are being given; and so a single series might not be heard by all. Then, too, many children have poor auditory imagery, which fact makes it difficult for them to reproduce material presented orally. However, in all of the tests a response is either entirely correct or it is entirely wrong. This "All or none" method of scoring is used in the writer's group tests, for the purpose of the series is not to discover how much better one child can perform a given task than another, but to see which children fail to perform it at all, i. e., to sort out the subnormal children who need to be given the individual tests before a correct diagnosis can be made regarding their mentality. That the responses must permit of no variation in interpretation, due to the examiner's judgment, is undoubtedly as essential in group tests as in the individual tests.

Altogether, twenty-five tests have been selected which seem to fulfill the requirements mentioned above. These have been divided into groups of five, designated for convenience, as five, six, seven, eight and nine year tests. These have been given to 904 school children distributed according to age as shown in Table I.

TABLE I.

Age	*No. Cases*
6.....(i.e., 5 yrs. 6 mos.....6 yrs. 5 mos. incl.).....	160
7.....(i.e., 6 yrs. 6 mos.....7 yrs. 5 mos. incl.).....	225
8.....(i.e., 7 yrs. 6 mos.....8 yrs. 5 mos. incl.).....	224
9.....(i.e., 8 yrs. 6 mos.....9 yrs. 5 mos. incl.).....	175
10.....(i.e., 9 yrs. 6 mos.....10 yrs. 5 mos incl.).....	120
Total	904

Description of the Tests

Tests for Year V

V, 1. Formboard.

Materials. This is an adaptation of Goddard's[4] formboard. From heavy white cardboard the ten well-known forms are cut. Black paper is pasted over the back of the openings to make them stand out more clearly, and then another sheet of

[4] Goddard: *The Training School.* Vol. IX, No. 4, June, 1912, pp. 49-54.

white cardboard is pasted over all this to form the back. Ten little cards which exactly fit the ten forms, are placed in an envelope pasted on the back of the formcard.

Procedure. The examiner holds up a booklet in front of the group of children to be tested, and says:

> "Open these little books to this black and white card. In the envelope on the back you will find a number of little cards. Take these out of the envelope and put them in a pile on your desk. Now, there is one place in the black and white card into which each of these little cards will exactly fit,—like this. (Illustrate with square.) See if you can find the place for each of the others."

Help may be given any child who has difficulty in finding the place for each card the *first* time. When all the children have the cards properly placed, have them remove them and again place them in a pile on the desk. Then proceed as follows:

> "Now you know where each little card belongs. This time we are going to see who can put all the cards in their places *first*. Everyone wait until I say 'Go,' and then work as fast as you can until I say 'Stop.' Then put your hands in your laps AT ONCE, without touching a single card again. Is everybody ready? Go."

Allow 50 seconds before giving the signal to stop. Care must be taken to see that no child touches a card after he has been told to stop.

Scoring. The test is passed if all ten cards are in their proper places at the end of 50 seconds. A plus or minus, indicating whether the child has passed or not, can be quickly marked on the blank page opposite the form card, by the examiner, before having the children replace the cards in the envelope.

V, 2. Counting 4 Circles. (Lowell)

Materials. A card, 6 in. x 18 in., on which are printed four large circles about 2 inches apart, is necessary, and also two wooden blocks or other small objects. Space for the child to make the crosses is provided at the top of page 2 of the booklet.

Procedure. "When I say, 'Make a cross,' this is what I want you to make. (Make X on the blackboard.) If I should say 'Make a cross for each of these blocks (holding up two blocks), how many crosses would you make? Yes, two,—one for this block, and one for that (XX), see?'

"On this large card are some circles. When I hold the card so you can see them, you are to count the circles TO YOURSELVES, and then make as many crosses up here at the top of your paper (Indicate) as there are circles. Ready. Look." Expose the circles for 10 seconds.

Scoring. The test is passed if four crosses are made.

V, 3. Copying Square. (Binet)[6]

Materials. Use a large cardboard on which an eight inch square is printed in heavy lines.

Procedure. Holding up the large square, and indicating the space in the booklet to be used, the examiner says:

"At this side of the big space, here, see how nicely you can make a square just like this one."

Hold the card so all can see it while they are drawing, and then when they have finished one, say:

"Now see if you can make a still better one at this other side of the big space."

Scoring. The test is passed if one of the two squares drawn is as good as those on the score card used by Binet.

V, 4. Discriminating Colors. (Binet)

Since it is impossible to have children in a group test name colors, the following adaptation of Binet's color test was used. It necessarily makes a different test from the original, since the directions must be comprehended and followed exactly.

Materials. One-half inch squares of the red, yellow, blue and green, are pasted on page 2 of the booklet in the order named.

Procedure. "Look at the colors in the middle of this page. Listen carefully, and then do just as I ask you:

"1. Make a cross *in* the *yellow* square, like this.
"2. Draw a line *through* the *green* square, like this.
"3. Now make a cross *above* the *red* square, like this.
"4. Draw a line *under* the *blue* square, like this."

In order that the child shall understand what is wanted, draw a square on the board each time, and do what you tell the child to do. Give directions slowly, and repeat each.

Scoring. The test is passed if all four colors are correctly marked.

[6] *Année Psychologique* (1911) vol. 17, pp. 145-201.

V, 5. Irregular Tapping. (Pintner,[6] Kuhlmann.[7])

The present adaptation has been made from the form used by Dr. Kuhlmann in his 1917 Revision of the Binet Tests.

Material. Use a card, 6 in. x 18 in., on which are printed four squares, about two inches apart. At the bottom of page 2 in the booklet are three rows of squares similar to the one on the examiner's card.

Procedure. "At the bottom of the page you see three rows of squares. The first row looks like this one, doesn't it? (Point to card in hand.) Now I am going to tap some of these squares, like this. (Tap the first two squares with a ruler.) Watch me carefully, and when I've finished tapping, put a cross, like this, X, in just the squares that I tap. Ready. Watch." Then tap the squares in Series A in the order named, at the rate of one square per second, and allow time between series for the child to cross the squares tapped. Always tell the children which row of squares to use for each series.

A) 1-2-4
B) 1-3-4
C) 2-3-4.

Scoring. The test is passed if one of the three rows is correctly marked.

TESTS FOR YEAR VI

VI, 1. Aesthetic Comparison. (Binet)

Material. On each of the pages 3, 4 and 5 of the booklet are two pictures used by Binet for the comparison.

Procedure. "Look carefully at the two pictures on this page (3). Make a cross under the prettier one of the two." Give the same instructions for each of the pages 4 and 5.

Scoring. The test is passed if the prettier face is indicated in two of the three series.

VI, 2. Mutilated Pictures. (Binet)

Materials. The four pictures used by Binet are printed on four successive pages in the booklet, 6, 7, 8 and 9.

Procedure. "Look at this picture. (Page 6.) Part of the face is gone. Let us see what part it is. You see the eyes. Look, the eyes are there. You see the nose. Yes, the nose is there, and look, the chin is there. Now what part is gone?

[6] Pintner: A Scale of Performance Tests — 1917.
[7] Kuhlmann: The Measurement of Mental Development — A School Publication — Faribault — 1917.

(Children name mouth.) Yes, the mouth is gone. Now make a cross right where the mouth should be, to show that that is the part that is gone.

"Look at the next picture. Do not tell me this time what is gone, just make a cross to show me what part is missing.

"Make a cross to show what is gone from the next picture (Page 8.)

"And now show what is gone in the last picture." (Page 9.)

Scoring. The test is passed if the last three pictures are correctly marked.

VI, 3. Counting Irregular Taps. (Kuhlmann)

Materials. The examiner needs a wooden block or the blunt end of a pencil for tapping. Ruled blanks are provided the children on page 10 of the booklet.

Procedure. "Listen. I am going to tap on the table and see if you can count the number of taps. You count to yourselves, and give me the number when I ask you. Ready."

Tap 5 times, at the rate of one tap per second. See that the hand is screened, while tapping, by a large cardboard.

"How many times did I tap? Five, that is right. Now, instead of telling me how many times I tap, I want you to make a cross for each tap, when I get all through. I just tapped five times, so how many crosses shall I make to show it? Yes, five, like this: XXXXX.

"Sometimes I will stop tapping and then begin again. Don't let that fool you. You count only those you hear, and when I ask how many, you will make one cross for each tap, here, on this first line. (Indicate line each time.) Ready. Listen."

Tap the following series in order, at the rate of one tap per second, for each cross, and pausing a second for each dash:

1) XX-XX (4)
2) X-XX-XXX (6)
3) XXX-X (4)
4) XX-XXX-X (6)
5) X-XX-X-X (5)

Scoring. The test is passed if 3 out of the five series are correctly counted as indicated by the correct number of crosses.

VI, 4. Two Simultaneous Commands.

This test was suggested by Binet's test in which three simultaneous commands were used.

Materials. One row of circles and a row of squares of different sizes are provided for the children on page 10 of the booklet.

Procedure. "See these squares and circles. Listen carefully and see if you can remember what I am going to ask you to do with them, when I am all through telling you.

"Make a cross in the BIGGEST square.
Then draw a line in the FIRST circle.

"I will tell you once more, and then see if you can do both things. Listen:

"Make a cross in the BIGGEST square,
Draw a line in the FIRST circle."

Scoring. The test is passed if both directions are correctly followed.

VI, 5. Perception of Sound. (Lowell)

Material. A bean-bag, a baseball, a penny, a pencil, a key, and a wooden block are used. Have a table which can be concealed from the children and from which the above named objects can be dropped. A card picturing each of the objects with respective symbols to be drawn by the children, is tacked up in view of the children. See Fig. I. Space for six drawings is provided at the bottom of page 10 in the booklet.

Procedure. "What is this?"

Hold up each of the objects and have the children name it.

"Now look at this card (Fig. I), and tell me what each of these pictures is. (Have them name them all.)

"Now I am going to drop each one of these on the floor where you cannot see it, and I want you to listen to the noise each makes, and see if you can tell which of these I drop.

"If I should drop the bean-bag first, you would draw, in this first space, here, this. (Point to symbol at the side of the picture of the bean-bag); if I dropped the key next, you would make this X in the second space, here; if I dropped the penny next you would make this line in the third space, etc.

"Now shut your eyes, and see if you can tell from the sound which one of these I drop, and then look at the card to see what to draw. Then draw that in the first space. Ready. Listen."

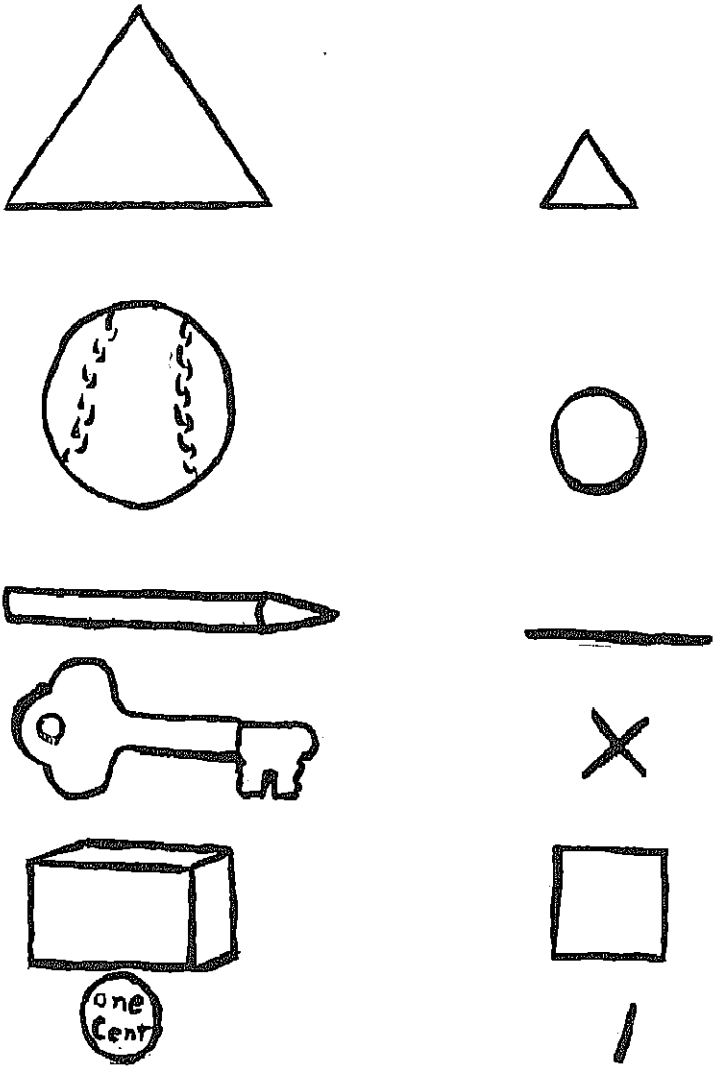


FIGURE I. (vi, 5)

Drop objects in the following order:

block
bean-bag
penny
key
ball
pencil

Scoring. The test is passed if five out of the six objects are named in the correct order.

TESTS FOR YEAR VII

VII, 1. Number of Toes.

Because it is impossible to prevent a large group of children from counting their fingers if one were to use the original Binet test, the number of toes on each foot and on both feet together, has been substituted.

Procedure. "I am going to ask you some questions, and I want you to write the answers, here at A)———, at B)——— and at C)——— Ready:

A) How many toes on your left foot?
B) How many toes on your right foot?
C) How many toes altogether on both feet?"

Scoring. Since most children of seven years can make figures up to 10, further use of crosses in answering questions is not permitted. The test is passed if the answers to all three questions are correct.

VII, 2. Memory Span. (Binet)

Procedure. "I am going to read you some numbers. Listen carefully, and see if you can write them when I get through just the way I read them."

Read the digits at the rate of one digit per second, without rhythm.

"Listen:

A) 6-5-2-8-1
B) 4-9-3-7-5
C) 2-8-6-1-9"

Scoring. The test is passed if the digits in one series are correct, even though their order be changed.

VII, 3. Copy Diamond. (Binet)

Material. Use a card on which a large diamond, 9 in. x 18 in., has been drawn in heavy lines.

Procedure. "At the left side of this big space, see how nicely you can draw a diamond just like this one."

Hold the card so that all can see it as they draw. After they have finished drawing the first diamond, say:

"Now draw another one at the right side of the big space just like this one."

Scoring. The test is passed if one of the two drawings is as good as those on the score card used by Binet.

VII, 4. Tying Bowknot. (Binet)

Material. Two pieces of tape, each about seven inches in length, are fastened to a stiff card pasted at the bottom of page 11, in the booklet.

Procedure. "You know what kind of a knot this is, don't you? (Show them bowknot already tied). Yes, it is a bowknot. I want you to tie the same kind of a knot with the pieces of tape you have, just as quickly as possible."

Scoring. The test is passed if a bowknot is correctly tied.

VII, 5. Geometrical Figures. (Lowell)

Material. Use five cards, on which are printed, in a different order, on each card, a circle, a square, a cross, and a triangle. See Fig. II.

Procedure. "I shall show you some cards for ten seconds each. On each card you will find these forms. (Draw on blackboard.) But the order in each case will be different. On one card a square will come first, and on another card the circle will come first, etc. You will notice carefully the order of the forms, and when I take the card away, you will draw what you saw in just the order it was on the card. Draw as quickly as possible. Ready. Look at the first card."

Allow no more than one minute for reproducing a card.

Be sure all children are giving attention when you hold up a card.

Scoring. The test is passed if three of the five series are correctly drawn.

TESTS FOR GROUP VIII

VIII, 1. Ball and Field. (Terman)

Material. The incomplete circle used by Terman to represent the baseball field, is printed in the booklet.

Procedure. "Let us suppose that your baseball has been

lost somewhere in this round field. You have no idea what part of the field it is in. All you know is that the ball is lost somewhere in this field. You are to mark out a path to show me how you would hunt for the ball so as to be sure not to miss it wherever it might be. If you draw just a line to the center from the gate, like this,—it would mean that you only looked that far and then stopped,—and the ball might be away over at the other side; you don't know where it is. Now

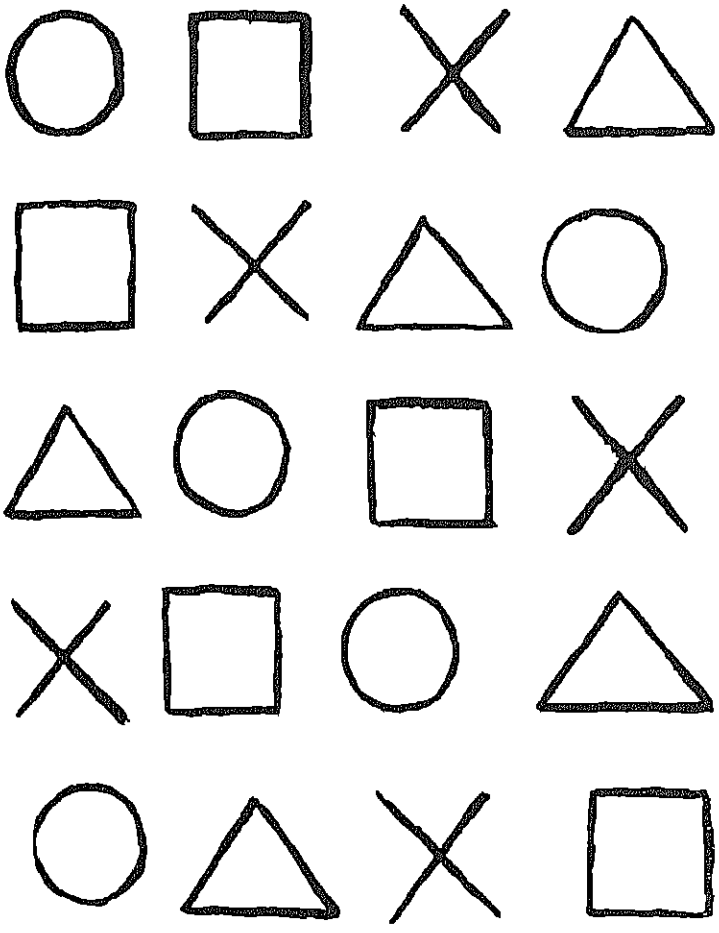


FIGURE II. (vii, 5)

begin at the gate, and with your pencil mark out a path to show how you would hunt for the ball so as to be sure not to miss it."

In giving an individual child this test, according to Dr. Terman's directions, the examiner watches what the child draws, and if he stops after drawing just a short path, the examiner asks him where he would go next if he hadn't yet found the ball. As such procedure is obviously impossible with a large group of children, the illustration is used in the directions.

Scoring. The test is passed if the child shows a plan as good as those shown on Terman's score card for eight-year-old children.

VIII, 2. Writing from Dictation. (Binet)

Procedure. "I want you to write something for me as nicely as you can on this long line. Write these words: 'See the little boy.' Be sure to write it all: 'See the little boy.'"

Scoring. The test is passed if the sentence is written legibly enough to be easily read, and if no words are omitted. Unless a word is so incorrectly spelled as to make it unrecognizable, the error is not counted.

VIII, 3. Counting the Value of Stamps. (Binet)

Material. Use a cardboard on which have been pasted three one-cent stamps in a row at the top, and three two-cent stamps in a row under these.

Procedure. "Write the answer to this question on the short line here." Hold the card before the children and say: "How much will it cost to buy all these stamps? A green stamp is worth one cent, and a red stamp is worth two cents. Now, how much are they worth altogether?"

Scoring. Test is passed if the answer, 9 cents, is given.

VIII, 4. Comprehension Test. (Kuhlmann)

Material. Three one-inch squares are printed in the booklet on page 3.

Procedure. Point to the parts as you mention them.

"See this first square. This is the *center* of the square. What is it? This is the *upper right corner* of the square. What is it? This is the *lower left corner.* What is it? This is the middle of the *left side of the square.* What is it?

"Now take your pencils, and draw a straight line from the center of the first (second and third) square to:

1) The upper left corner.

2) The middle of the upper side.
3) The lower right corner of the square."

Scoring. The test is passed if two of the three lines are correctly drawn.

VIII, 5. Estimation of length of lines. (Kuhlmann)

Material. On separate pages of the booklet are drawn three lines. The first line, drawn vertically in the center of the page, is three inches long; the second line, similarly drawn on another page, is two and one-fourth inches long; and the third line, drawn horizontally, and having two curves, is three inches long.

Procedure. "I want to see how well you can judge the length of lines without a ruler or even your pencils to help you measure. Look at this line carefully, and then draw a line under it here, horizontally, i. e., from left to right like this, (hold pencil where line is to be drawn), which is

1) Just as long and no longer than this line.

"Now look at this next line. Draw a line under it, here, horizontally, which is:

2) Just twice as long as this one.

3) "Let's play that this is a piece of string. If I could take hold of the ends of it, and pull the string out straight, how long a piece would it be? Draw a straight line here, to show how long it would be." (Point to space under the curved line.)

Scoring. The test is passed if two of the three lines are drawn within one-half inch of the required length, i. e

1) 3 in.
2) 4½ in.
3) 3 in.

TESTS FOR GROUP IX

IX, 1. Drinking Cup. (Bureau of Analysis, etc., N. Y.)[8]

Material. A square of white paper, 7½ x 7½ in.

Procedure. "I am going to show you how to make a drinking cup, and when I finish, I want you each to make one just like it. Watch carefully.

1st: Fold the square through the center like this.

2nd: Take the right hand corner and fold it over to the opposite side so that these two edges are equal.

[8] New York State Board of Charities, Bureau of Analysis and Investigation: Eleven Mental Tests Standardized, 1915, *Eugenics and Social Welfare Bulletin* No. V. p. 42.

3rd: Next, take the other corner on the left side, and fold it over the same way so that the edges come straight across.

4th: Then, fold down this flap, and put it into this outer space to hold it.

5th: Now fold down the top flap for a cover, and you see how the cup looks.

"You take your square of paper now, and make one."

Scoring. The test is passed if the five folds of the drinking cup are correctly made.

IX, 2. Maze. (Porteus)[9]

Material. The maze used by Porteus for nine year children, is printed in the booklet. See Fig. III.

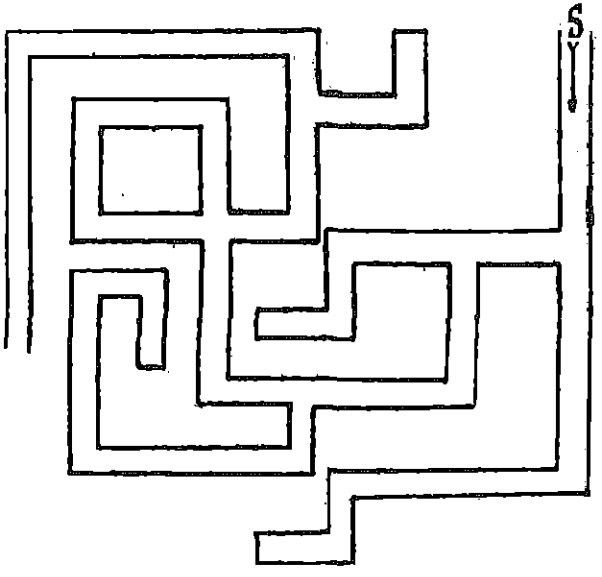


FIGURE III. (ix, 2)

Procedure. "Here you have a maze. Some of these paths are open, so you can get through, and some are closed at the ends. See how quickly you can find the shortest way out, without getting into any of the closed paths. Start where you

[9] Porteus: Mental Tests for Feeble Minded; *Journal of Psycho-Asthenics*, Vol. 19, No. 4, June, 1915, pp. 200-213.

see the letter *S*, and mark out the very shortest path you can find in order to get out over here on the other side. Do not stop to erase anything. If you make a mistake, just draw a line across it and go on as fast as you can."

Scoring. The test is passed if the shortest path is marked out in one minute, with not more than two errors. An error consists of a line, however short, drawn out from the shortest path.

IX, 3. Tapping Squares. (Whipple)[10]

Material. A sheet of paper is used on which a rectangle has been printed. This rectangle is divided into 150 half-inch squares, arranged in ten rows of fifteen squares each.

Procedure. "Please take your pencils and hold them firmly about half way up. Place your arm comfortably on the desk so that it may bend freely at the elbow. Now when I say, 'Ready,' you are to tap as many squares as you can, in order, without hitting any square twice, or missing a square, or touching any of the lines. Work until I say 'Stop.'"

Scoring. The test is passed if 50 squares are correctly tapped in the 30 seconds allowed.

IX, 4. Alphabet Test. (Kuhlmann)

Procedure. "Of course you all know the letters of the alphabet, a-b-c-d-e etc. Who can tell me what is the first letter before 'C'? (B). Now who can tell me what is the sixth letter before 'X'? (R). Let us put this one on the board, R-S-T-U-V-W-*X*. Notice that W would be the first letter before X, V the second, U the third, T the fourth, S the fifth and R the sixth.

"I will give you 30 seconds to answer each one of these questions.

1) What is the 3rd letter before K (H)
2) " " " 5th " " W (R)
3) " " " 2nd " " F (D)
4) " " " 1st " " Q (P)
5) " " " 4th " " E (A)"

Always write the letter to be used as a basis on the board, so that the children will not misunderstand.

Scoring. The test is passed if three of the five letters are correct.

[10] Whipple: *Manual of Mental and Physical Tests*, 1910. pp. 100-115.

IX, 5. Memory Test. (Kuhlmann)

Procedure. "I am going to write five words on the board, and let you look at them for a few seconds. Then I will erase them and write five others and let you look at them, and so on with the other series. When I am through, I will give you the first word of a series, and you are to write down from memory, the other four words that go with it. Then I will give you the first word of the next list, and you will write the four that go with that one, etc. You must not write down any until I tell you to."

Let them look at each series for 30 seconds, then erase and take the next. Give them 30 seconds to recall each list. The lists are:

Silver	red	kitchen	blackboard	lawn
spoon	cow	stove	teacher	grass
table	fence	flour	desk	tree
bread	clover	water	slate	house
fork	dog	spoon	lesson	walk

Scoring. The test is passed if the child gives three series without more than one error in each. An omission counts as an error.

Difficulties of Administering Group Tests

Many problems, which from their very nature can never be entirely eliminated, arise in giving group tests to the primary grades. Some of these difficulties are due to the distractibility of young children, to their inability to comprehend directions readily, to their undeveloped moral sense, to the attitude of the teacher, and to the apparent inhibition of response which the nervous child exhibits.

The inability of the young child to attend to set problems for any length of time is a serious difficulty. It necessitates devoting almost the entire first year of his school life to the formation of habits of work, of play, of thought and of living. Up to the time of his entrance into school, he is permitted to eat, sleep, and play largely at will, but with the beginning of his educational career he is expected to live according to a more or less definite plan, regardless of his own wishes or pleasures. When a stranger tries to test forty such children, before regular habits have been formed, the difficulties encountered may easily be imagined. Usually, the children are sufficiently shy and timid not to need disciplining, but sometimes that very timidity prevents them from being responsive at first. Therefore, their confidence must be won by the time

the booklets are distributed and the examiner is ready to begin the testing. The fact that the children are not permitted to look through the pages except as they use them for the tests helps to keep up the interest. When the booklets are finally distributed, the directions for the first test given, the children attentively (?) awaiting the signal to begin work, and then a little voice in the corner confidentially informs one that "My sister has some gold beads just like yours," the examiner wonders if he can ever be sure of a child's attention.

The difficulty experienced by young children in comprehending directions has been discussed under the "Selection of Tests." Frequent illustrations, full and explicit directions with numerous repetitions, all help the child in understanding what he is to do. However, ability to comprehend and follow directions accurately, is a test of intelligence, and too many or too great a variety of instructions may destroy the value of the test.

In order to stimulate the child to his best effort, one appeals to his competitive instinct, his sense of rivalry. This has its disadvantages, for the child in his eagerness to do better than his neighbor, or to get a word of praise from the examiner, resorts to methods "fair or foul." It is generally agreed that the young child is non-moral. The disapproval of the group has not been experienced sufficiently to make a lasting impression; and so the child slyly copies from his neighbor, or covertly writes on his desk the digits which are being read. To guard against such conditions as these, the writer has used various devices. Some have been mentioned under *Procedure*, such as having the child place his hands in his lap the instant he is given the signal to 'Stop;' cover his work with his hands as soon as he has finished the task assigned; or, in the digits test, place the pencils on the desks and leave them there until told to write. In one or two instances, where the attention had several times been called, without effect, to the fact that someone in the room was not playing fair, more drastic measures have been taken, such as speaking to the child by name, or changing his seat. However, as most children are very sensitive to criticism, such measures are seldom necessary.

Frequently a teacher, not knowing the object of the group testing, feels that unless the pupils respond well to the tests, her reputation as a teacher will be at stake. In her anxiety to have the children make a good showing, she offers suggestions which help them. For example, when trying out various tests for the nine-year group, the writer used Terman's tes

of finding all the words that rhyme with "Day, mill and spring." In one third grade, the teacher felt that the children were not responding as readily as they should, so with much irritation she spoke to the class. "Why, children, you know how we have been finding words to go in the 'ING' family, so I don't see why you can't find others like 'Day' to go in the 'AY' family." Of course the children could find words after that, but the test, as such, had lost its value. To avoid such incidents, the examiner should explain to the teacher at the outset that the purpose of the work is to see what each child can do by himself, without any suggestions from her.

So far, no difficulty has been experienced in securing the proper attitude on the part of the child, for the interest and novelty of the tests counterbalance any temperamental disturbances. However, it has been evident in the cases of several nervous children that the excitement produced by the competition, novelty and mental effort has inhibited responses, especially where the test has a time limit. If a pupil happens to be an only child and has not yet learned to adapt himself to an environment in which there are many children, he may fail to make an average score in the group tests, though when tested individually he may test even above normal.

Table of Norms and How to Use It.

If group tests are to be of value in selecting subnormal children, the number of tests passed by an individual must correlate with his age and intelligence. Table II shows the average number of tests passed by children of each chronological age and of the various I. Q.'s found in Grades I, II, and III.

As shown by the above table, the chronological ages of the children ranged from six to twelve years. By six years is meant all ages from five years six months to six years five months, inclusive; by seven years is meant all ages from six years six months to seven years five months, inclusive; etc. In order that the average number of tests passed by children of any age might serve as an index of their intelligence, individual I. Q.'s were obtained by means of Dr. Kuhlmann's 1917 Revision of the Binet tests, and then the average number of group tests passed for these various intelligence quotients and for each chronological age was computed. Thus, in using the table of norms in classifying an individual case,

TABLE II

NORMS. AVERAGE NUMBER TESTS PASSED AT VARIOUS AGES AND GRADES

Grade I. (Group Tests V—VI—VII.)

I. Q.	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.	11 yrs.
.70.....	2.0	4.6	6.2	8.3	....	....
.80.....	4.3	6.6	9.6	11.1	....	....
.90.....	6.1	8.4	11.8	....	....	....
1.00.....	7.7	11.2	14.5	....	....	....
1.10.....	8.2	12.6	....	....	....	....
1.20.....	10.5	14.0	....	....	....	....
1.30.....	13.0	15.0	....	....	....	....

Grade II. (Group Tests VI—VII—VIII.)

I. Q.						
.70.....	....	....	7.3	8.6	10.1	....
.80.....	....	....	8.8	11.0	12.0	....
.90.....	....	8.6	9.9	12.4	14.2	....
1.00.....	....	9.6	11.4	13.7	....	....
1.10.....	....	10.1	12.5	14.5	....	....
1.20.....	....	11.5	13.5	....	....	....
1.30.....	12.0	13.3	....	....	....	....
1.40.....	13.4	15.0	....	....	....	....

Grade III. (Group Tests VII—VIII—IX.)

I. Q.						
.70.....	....	....	....	8.0	9.8	10.3
.80.....	....	....	8.5	9.2	10.8	11.6
.90.....	....	....	9.5	10.2	11.9	....
1.00.....	....	....	10.0	12.0	13.0	....
1.10.....	....	....	12.0	13.5	14.2	....
1.20.....	....	....	13.1	14.6	....	....
1.30.....	....	12.0	14.2	....	....	....
1.40.....	....	14.0	....	....	....	....

one compares the number of group tests passed by him with the average number of tests passed by children of the same grade and chronological age, and finds the average in the series and its corresponding I. Q., with which his score agrees. For example, if an eight-year-old child in the first grade passed only seven tests, his score would lie between the average for children of his age who had an I. Q. of .70, namely 6.2 tests, and the average of those with I. Q. of .80, or 9.6 tests. This would give him a group I. Q. of .72. Since such a rating would place him in the sub-normal group, he should undoubtedly be given the individual Binet tests, in order that his mental status be accurately gauged.

Reliability of the Group Tests

In order to determine the reliability of any series of tests, the data must be considered from various points of view. The aspects which will be discussed in connection with the reliability of the present scale are; *first*, the discriminative capacity of the tests, i. e., their ability to distinguish between children of various ages, as shown by the increase in the per cent passing the tests for successive chronological ages; *second*, the percentage of children passing "at age," i. e., the percentage which will cause average six-year-olds to show a mental age of six years, average seven-year-olds to show a mental age of seven years, etc., as compared with the percentage passing at age in the individual tests; *third*, the ability of the group tests to discriminate between various mental ages in three school grades where the same group of tests has been given; *fourth*, the correlation between group and individual tests, as calculated by Karl Pearson's method; *fifth*, the agreement between group test results and school attainment; and *sixth*, the correlation between the writer's scale and an "Absolute Point Scale," [11] when used on the same children.

The Increase in Per Cent Passing for Successive Chronological Ages

The discriminative capacity of group, as well as of individual tests, is measured by the increase in the percentage of children passing the tests from one age to the next. The greater the increase in the percentage passing any test from one age to the next, the higher is the discriminative capacity of that test.

Table III compares first, the percentage passing at age for each test in the five age groups, with the percentage of children, chronologically a year older, passing each test; and second, it compares the per cent increase from one chronological age to the next as found in the group tests, with the per cent increase found by Kuhlmann [12] and Terman,[13] where they have used the same tests in their revisions as individual tests. For example, 66 per cent of the six-year-old children passed the aesthetic comparison test (VI,1) when given in the group, and 75 per cent of the seven-year-olds passed it, thus giving

[11] G. Arthur and H. Woodrow: An Absolute Point Scale. *J. Appl. Psychol. Vol.* 3, 1919, pp. 118-137.

[12] These figures were taken from results in Dr. Kuhlmann's 1912 revision, in which the tests had not been correctly placed in the scale.

[13] Terman's results as quoted are taken from the data published in 1917 on the Stanford Revision.

an increase from one age to the next of 9 per cent. When Kuhlmann used it as an individual test, he found an increase of 4 per cent passing from ages six to seven, and Terman found an increase of 6 per cent at the same ages.

TABLE III

INCREASE IN PER CENT PASSING FOR SUCCESSIVE CHRONOLOGICAL AGES

Group V	Yr. 6 (121 cases)	Yr. 7 (135 cases)	% Increase	Individual Kuhlmann	% Increase Terman
1. Formboard	65	84	19	..	..
2. Counting 4 Circles	66	87	21	1	..
3. Copying square	85	95	10	..	..
4. Discriminating Colors	72	81	9	..	9
5. Irregular Tapping	70	85	15	..	..

Group VI.	Yr. 6 (122 cases)	Yr. 7 (136 cases)	% Increase	Individual Kuhlmann	% Increase Terman
1. Aesthetic Comparison	66	75	9	4	6
2. Mutilated Pictures	63	74	11	6	15
3. Irregular Taps.	60	88	28	..	..
4. Two Commands	54	75	21	..	..
5. Perception of Sound	51	74	23	..	..

Group VII.	Yr. 7 (95 cases)	Yr. 8 (198 cases)	% Increase	Individual Kuhlmann	% Increase Terman
1. No. of Toes	70	88	18	17	9
2. Memory Span.	63	81	18	3	6
3. Copy Diamond	84	92	8	11	18
4. Tying Bowknot	72	81	9	..	16
5. Geom. Figures.	61	76	15	..	..

Group VIII.	Yr. 8 (199 cases)	Yr. 9 (174 cases)	% Increase	Individual Kuhlmann	% Increase Terman
1. Ball and Field.	70	79	9	..	7
2. Dictation	78	87	9	..	8
3. Value of Stamps	70	84	14	20	17
4. Comprehension Test	72	81	9	..	..
5. Length of Lines	69	81	12	..	..

Group IX.	Yr. 9 (106cases)	Yr. 10 (112cases)	% Increase	Individual Kuhlmann	% Increase Terman
1. Drinking Cup..	67	78	11	..	..
2. Maze.....	73	83	10	..	..
3. Tapping Squares.....	70	79	9	..	..
4. Alphabet Test.	73	83	10	..	..
5. Memory Test..	60	71	11	..	..

Unfortunately, there are no other similar group tests published, with which the writer's percentages passing the various tests can be compared. However, there is about as much variation existing between the data from the two revisions of the Binet Scale as between the individual and group results.

Per Cent Passing "at Age" in Individual and Group Tests

Considerable difference exists between the percentage of children passing "at age" in the group and in the individual tests. Kuhlmann has found that a much higher percentage of young children pass "at age" than of the older ones; thus, he found that 88 per cent of the four-year-old children pass the four year tests, whereas only 54 per cent of the twelve-year-olds pass the twelve year tests. However, in the group tests, the writer finds as small a percentage of six and seven year children passing six and seven year tests, as of nine-year-old children passing the nine year tests. Table IV compares the percentages passing "at age" as found by Kuhlman[14] in the Individual Tests with those by the writer in the Group Tests. Thus, 78 per cent of the six-year-old children pass the six year Individual Tests, and only 60 per cent of them pass the six year group tests.

TABLE IV

Per Cent Passing At Age in Individual and Group Tests

C. A.	6 Yrs.	7 Yrs.	8 Yrs.	9 Yrs
Kuhlmann—Individual Tests.	78	73	68	64
Lowell—Group Tests.....	60	69	72	68

The explanation of these differences in the six and seven year results may be found in the fact that a group test is largely a matter of comprehending and following directions

[14] These figures were given me through the courtesy of Dr. Kuhlmann, from his manuscript of *"The Measurement of Mental Development, a further Extension and Revision of the Binet-Simon Tests"* to be published soon.

presented orally, without any individual help for individual needs. If a child has not been accustomed to doing things in a group, he will be easily distracted. Then, too, the younger the child the less capable he is of following instructions; but with each additional year of experience, and with an increasing mental age, the performance of tasks become easier. The effect of these two factors, namely, increasing mental age and increasing number of years in school, on the number of group tests passed, is shown in Table V.

TABLE V

RELATION OF M. A., NO. YEARS IN SCHOOL AND AV. NO. OF VII YEAR GROUP TESTS PASSED

M. A.	Average Number of Tests Passed in: Grade I	Grade II	Grade III
4— 4.9 years	0	...	...
5— 5.9 years	0.7	1.0	...
6— 6.9 years	1.2	3.0	3.3
7— 7.9 years	2.7	3.4	4.4
8— 8.9 years	4.0	4.3	4.5
9— 9.9 years	...	4.5	4.5
10—10.9 years	...	...	4.5

The data for the above table was obtained by testing 277 children individually, with Kuhlmann's 1917 Revision,—thus getting their mental ages. Then the average number of tests in the VII year group—this group was the only one of the three which was given in all three grades—passed by children of the various mental ages, was found for each grade. Thus, children with a mental age of 6—6.9 years, who have been in school less than one year,—that is, who are in the first grade, on the average pass 1.2 tests of the VII year group; children of the same mental age who have been in school nearly two years, or are in the second grade, pass 3 tests; and those in the third grade, of the same mental age, pass 3.3 tests. So, with each additional year in school the number of tests passed increases, even though the mental age remains constant. On the other hand, with the "time in school" factor constant, but with increasing mental age, one also finds an increase in the number of tests passed. Thus, in grade I children whose mental age is 4—4.9 years will not pass any of VII year tests; those whose mental age is 5—5.9 will average 0.7 of a test; whereas those having a mental age of 8—8.9 will average 4.0 tests.

These facts seem to indicate that two important principles are involved in the group responses of primary children.

First, that the distractibility, so noticeable in kindergarten and first grade children, gradually decreases with each additional year of group or school performance, until it becomes, in the third grade, a negligible factor. In other words, the individual third-grade child has had time to adjust himself to his new environment, and to realize through frequent encounters with other members of the group, that his wishes and his behavior must conform to theirs. Second, mental age is an important consideration. This is true, however, only in so far as the work is sufficiently difficult to require effort on the part of the individual. Thus, in the third grade, taking carelessness and chance error into account, the maximal performance is reached, for the VII year tests, at a mental age of eight. Further increase in mental age does not improve the average eight year score.

Correlation Between Group and Individual I. Q.'s.

Since the standardization of the group tests has been completed, the writer has given the group tests to 153 children in the primary grades and found their group I. Q.'s. Then all of these children were tested individually by Kuhlmann's 1917 Revision of the Binet tests, and their individual I. Q.'s worked out. The coefficient of correlation, computed by Karl Pearson's method, for the two series of I. Q.'s is .77.

Group and Individual I. Q.'s vs. School Attainment

Individual records in group tests form a basis for promotion in primary grades, for the results obtained by group tests agree more closely with school attainment than do the results obtained by individual Binet testing. As a rule, children who give poor response to group tests, do poor school work, while those who obtain group I.Q.'s above 1.10 are capable of special promotion.

On the basis of Terman's [15] classification of normal or average children, namely, those having I.Q.'s from .90 to 1.10, the 153 children mentioned above were grouped into three classes. Those having individual I.Q.'s below .90 were called dull, those with I.Q.'s from .90 to 1.10, inclusive, were considered average or normal, and those with I.Q.'s above 1.10 were classed as superior. Then these cases were redistributed, on the basis of group I.Q.'s, into the same three classes. A comparison of the results is given in Table VI.

[15] L. M. Terman: *The Measurement of Intelligence*, 1916, p. 94.

Table VI

CLASSIFICATION OF 153 CASES BY GROUP AND INDIVIDUAL I. Q'S.

By:	Dull (Below .90) No. Cases	Average (.90 to 1.10) No. Cases	Superior (1.10+) No. Cases
Individual I. Q.'s....	27	80	46
Group I. Q.'s.....	49	66	38

By this table, one notes that 22 more cases are rated as dull by the group scores than by the Binet I.Q.'s. School records of these 22 children offer the following bits of information:

Case 1. RR.[10]
Case 2. Very poor school work, Marks average 70-76.
Case 3. R.
Case 4. RR.
Case 5. Deaf,—hence class work unsatisfactory.
Case 6. RRR.
Case 7. School work just passing—even with aid of special help.
Case 8. R.
Case 9. Conditioned in grade. Neuropathic heredity. Does nothing in group, and only with much coaxing can information be elicited when the child is alone.
Case 10. R.
Case 11. R.
Case 12. Father of child f.m.—one brother an idiot—an aunt an idiot.
Case 13. R.
Case 14. R.
Case 15. Has very poor comprehension. Has to have much individual help.
Case 16. R.
Case 17. R.
Case 18. RR.
Case 19. Stammers very badly.
Case 20. Conditioned. Can only make grade with individual help.
Case 21. R.
Case 22. R.

[10] R = repeated work of grade.
RR = repeated twice the work of the grade.
RRR = repeated three times the work of the grade.

Of the 66 cases classed as normal by the group tests, 53 receive a similar rating by the individual tests, and the other 13 were classed as superior. One must remember in this connection that a difference in an I.Q. of one point may change a child's rating from dull to normal or from normal to superior. For instance, a group I.Q. of 1.10 will place a child in the normal group, whereas an individual I.Q. of 1.11 would classify him as superior, and vice versa. A difference in classification due to a difference of 3 points or less between group and individual I.Q.'s occurred in 5 of the 13 cases mentioned above. The school records of the other eight give the following comments:

Case 1. Slow to respond. Never volunteers to recite in class.

Case 2. Schoolwork inferior to Individual I. Q. Does poor team work.

Case 3. Lazy and careless — will not work if possible to avoid it.

Case 4. Doesn't work well in group — wants attention centered on self.

Case 5. Has been in school less than one year. Spent one day in each of Grades I and II and then was located in grade 3B.

Case 6. R.

Case 7. Can't adjust to conditions in a group either at work or play. Is irritable and nervous. Mother insane.

Case 8. R.

Agreement is very close between Group and Individual ratings of the superior group. One interesting case is found on school records for this class, namely that of a child who was doing such good school work that the teacher recommended a special promotion for him, although his I. Q. as determined by the Binet tests did not warrant it. In view of his excellent school work, and of the fact that when given the group tests his score placed him in the superior group, the child was given the extra promotion. He has been in the new grade nearly two months and is still doing superior work.

Careful consideration of available school records seems to indicate and in fact, to emphasize a closer agreement between group test results and school attainment than between Binet scores and school work. Such a correspondence is to be expected, however, when one realizes that the principles involved both in group testing and in school work are fundamentally the same.

CORRELATION BETWEEN TWO GROUP INTELLIGENCE SCALES

While the present writer has been devising and standardizing a group intelligence scale by the "all-or-none" method, an "Absolute Intelligence Scale"[17] has been constructed in which a point method has been used. Nine well known types of tests have been used, namely:

1) Immediate Memory Span
2) Easy Opposites
3) Hard Opposites
4) Substitution
5) Word Building
6) Language Completion
7) Anagrams
8) Cancellation
9) Comprehension

Without discussing the mathematical calculation of their points, a fairly accurate idea of the general plan of the scale may be obtained from the following quotation:

"The quotients which we have termed discriminative values represent an absolute increase in ability from one age to the next, for their respective tests. ———— The final absolute intelligence scale is ———— secured simply by adding together the point scales of the individual tests. The normal absolute intelligence for any age is simply the sum of the normal point scores of that age for all nine tests."

This point scale differs from the all-or-none scale in practically every detail except that both are Group Intelligence Scales, and both can express their scores in terms of intelligence quotients. The two series of tests were given to a group of 77 children, and I. Q.'s for each scale computed. The coefficient of correlation between the two groups of I. Q.'s is 0.76 (Karl Pearson.)

SUMMARY

The present group scale offers a means for measuring the intelligence of large groups of children accurately enough to sort out all children of questionable normality. It also forms a basis for school promotion and demotion in the primary grades.

It consists of 25 tests, arranged in groups of five, and for convenience, called five, six, seven, eight and nine year tests.

[17] Grace Arthur and Herbert Woodrow: An Absolute Intelligence Scale. The writer is indebted to Miss Arthur and Dr. Woodrow for their cooperation in comparing group results and for making their data available for the writer's purposes.

The tests have been standardized on the basis of results obtained from 904 children of the ages five years six months to ten years and five months, inclusive.

The uncorrected coefficient of correlation (Pearson) between measurements made, in terms of intelligence quotients, by the Group Scale and by Kuhlmann's 1917 Revision of the Binet Tests is 0.77.

The group results form a more accurate index of the child's performance, than do the individual results obtained by the Binet tests.

The uncorrected coefficient of correlation between two series of I. Q.'s obtained for 77 children by the present group scale and by "An Absolute Point Scale" is 0.76.

A STANDARDIZED TEST FOR OFFICE CLERKS

By L. L. THURSTONE, Division of Applied Psychology, Carnegie Institute of Technology, Pittsburgh, Pa.

The clerical examination here reported has been taken by about five thousand office clerks and has been standardized by the Phoenix Mutual Life Insurance Co., the Equitable Life Assurance Society, the Westinghouse Electric & Mfg. Co., the Bell Telephone Co., Western Electric Co., Strawbridge & Clothier, Armstrong Cork Co., and others.

In devising a vocational test the psychologist should be guided by three fundamental considerations affecting the content of the test. 1) The difficulty of the content should correspond to the intelligence level of acceptable candidates. 2) The content should have an interest appeal to candidates who would be interested in the work tested for. 3) The special abilities should be represented in the test if there are any demonstrable special abilities. I do not believe that office work has any special abilities that have so far been demonstrated and hence I have confined myself to the first two criteria; namely, an appropriate intelligence level and content that appeals to the applicant for an office position. This reduces itself to the same type of problem that we find so frequently in preparing vocational tests; namely, the preparation of an intelligence test out of relevant content.

In deciding on the length of a vocational test we may be guided by two diametrically opposed considerations. If we have in mind the convenience of the interviewer or examiner in the employment office we should be tempted to use a short ten-minute test. So many distracting influences may appear in ten minutes and so much depends on the candidates "warming up" to the test that I should hesitate rejecting an applicant on what he does in a ten-minute interval. However, a ten-minute test is diagnostically more valuable than a ten-minute interview and if the two are combined, so much the better. The opposite consideration is that if no test is given, the applicant is put to work under more or less informal observation, and if the supervision is at all adequate it will appear during the course of a week or two whether the new employee is unusually capable. But two or three weeks repre-

sent many hours, not minutes. Now, why should we not give the employee a *standardized sample office job* which requires about 45 to 60 minutes? By means of a sample office job which has been standardized we may know more about the candidate's capacity for a variety of office jobs than can possibly be judged from a short list of opposites or other brief general intelligence test. These are some of the assumptions on which the clerical examination was devised.

The following is a brief description of the eight parts of the examination.

A. In this part the candidate checks the errors in addition and subtraction. There are only a few errors on the page of 120 additions. The small number of errors is intended to duplicate actual office conditions in which errors are even at worst more infrequent than correct answers. The careless candidate sees so many of the additions correctly performed that he does not scrutinize each item sufficiently to discover the errors.

B. In this part the candidate reads a section of Arnold Bennett's "Mental Efficiency" in which about forty of the words are incorrectly spelled. He is instructed to underscore every word that is incorrectly spelled, but the careless candidate becomes interested in the content and forgets to look for spelling, a condition which is typical in many office situations.

C. In this part of the examination the candidate cancels the four letters X, Z, U, and C. This combination of letters is rather difficult to keep in mind and taxes the attention of the candidate.

D. This is a short code learning test in which the systematic worker can profit from attending to the task in hand and thereby soon learning the letter-digit combinations. By doing so he gains time.

E. This is an alphabetizing test in which the candidate writes a list of forty names, placing each name in one of ten spaces and alphabetizing each of the ten groups of names. This test shows not only handwriting but affords a good opportunity for systematizing a simple clerical job since the candidate can choose among many different methods of doing the task.

F. This is also a test in which the candidate can profit by planning the task. He is expected to designate a list of insurance policies in three classes, according to the kind of policy, the date of issue, and the amount. He must keep these three attributes of the policies in mind simultaneously or so arrange his task that he can deal with each attribute sepa-

rately. The method of performing the classification is necessarily chosen by the applicant.

G. The Clerical Examination would not be complete without an arithmetic test. This part of the test contains twelve simple problems in arithmetic, including addition, subtraction, multiplication, fractions, and percentage.

H. The last part of the Clerical Examination is a bona fide general intelligence test and consists in matching ten proverbs with ten other proverbs so that the two proverbs in each pair have the same meaning or moral.

The Clerical Examination is given as an omnibus test with a maximum time limit of 90 minutes for the whole pamphlet. The average time is approximately 40 minutes. The instructions to the examiner are as follows:

"Give the applicant a copy of the Clerical Examination and ask him to do what it tells him to do. Select for him a quiet table where he may work without being disturbed. Record on the cover of the applicant's paper the 'Starting Time.' Tell him to return the blank to you or your assistant when he has finished. When the applicant returns the papers, record his 'Finishing Time.'"

I have found it convenient to eliminate the test proper from the first page of mental test blanks. The test proper is begun after the candidate turns the title page. This feature gives the examiner better control over the time, which is especially important with tests of short duration. On the title page of the Clerical Examination I have arranged a boxed space for a complete record of performance so that the examiner has a definite square in which to record each of the various scores, times, and percentiles. The headings in the boxed space are Finishing Time, Starting Time, Total Time, Speed Rank, Test A, Test B, Test C, Test D, Test E, Test F, Test G, Test H, Accuracy Score, Accuracy Rank, and Combined Rank.

When these pamphlets are filed with the employee's personnel records his performance in the Clerical Examination are easily referred to since it is summarized on the title page. The Accuracy Score is the sum of the scores allotted to each of the ten tests. The Accuracy Rank is the percentile rank of the Accuracy Score. The Speed Rank is the percentile rank of the Total Time.

Statistical Evaluation

In order to ascertain the diagnostic value of the examination it was given to one hundred employees of a large insurance company ranging from minor executives to young office clerks doing routine work. The employees were rated in five classes

according to the grade of office work in which they are employed. In addition to this information the age and schooling of each candidate was ascertained for correlation purposes. The following results are quite interesting.

Grade of office work actually being done by the candidate correlates with:

Accuracy in test	r +.50
Speed in test	r +.42
Schooling	r +.47
Age	r +.35

By multiple correlation the following coefficients have been ascertained between the grade of office work being done by the candidate and:

Accuracy and Speed combined	r +.61
Schooling and Age combined	r +.52
Accuracy, Speed, Schooling	r +.64
Accuracy, Speed, Schooling, Age	r +.67

It is significant, as shown by these correlation coefficients, that speed and accuracy in the Clerical Examination give a more reliable prediction of ability in office work than age and schooling combined.

In evaluating speed and accuracy by the method of multiple correlation it so happens that the two regression coefficients are alike, which means that in the combined score each minute of time is equal in weight to one of the 100 points of accuracy. This fortunate coincidence facilitates the determination of a combined score which represents both speed and accuracy.

Owing to the extensive use of the test it is not advisable to publish the detailed test which would make it available for candidates who might be scheduled to take the examination. Therefore, I have been content to describe the nature of the test and especially the statistical findings. However, I shall gladly send samples of the test to psychologists and other responsible parties who are interested in the investigation.

THE LEARNING CURVE IN TYPEWRITING

By J. Crosby Chapman, Western Reserve University

Object of the Study

The adequate study of improvement in mental functions which concern the educational process must satisfy the following three requirements: (1) the measurement must involve a group of pupils of reasonable size; (2) the study must extend over a prolonged period of time; (3) the investigation must, as far as possible, be incidental to the school work. The measurement of a group is essential if safe general laws are to result; extended practice is necessary because the curve of learning is by no means linear; practise should be under school conditions, otherwise the norms of performance cannot be applied to ordinary school work. Fortunately such a study is possible; typewriting as taught in the various business schools and high schools of commerce furnishes conditions, which while in no way equal to the exactness of laboratory experiments are yet sufficiently under control to be serviceable. Experiments on large groups found in such schools, working for long periods under normal circumstances, satisfy the conditions that have been laid down and give evidence concerning the general nature of changes in the rate of improvement in this complex function.

The present research was undertaken with a double object: (1) to obtain evidence with a view to a systematic study of the psychology of skill from the objective standpoint. (2) To investigate the rate of improvement of the subjects, in this particular school, using this particular method of instruction, in order to establish norms for the comparison of the relative values of different methods of instruction. The latter investigation has interest from the administrative standpoint; in the commercial schools of the country, various methods are employed and it is a matter of some importance to see which method yields the most satisfactory results.

Method of the Study

The general method of the research was to test, each week in the school year, the speed of a mixed class, ages 16-18, in typewriting. The object of the weekly test was explained, each pupil being given a copy of the following:

Information concerning the weekly test.

The object of these tests, which will be given once a week,

is to measure the improvement of each pupil, and the improvement of the school, in speed of typewriting. Each week you will be given a five minute test, consisting of fresh material which is unknown to you. Your object is to type as much of this material as you can in the five minute period, remembering that errors count against you, to the extent of one word off for each error. If you wish to get the maximum score, you will have to decide, as you would in a game, whether it pays you to go quickly and make mistakes, or whether it pays to go slowly and not make mistakes. You alone can decide this. Your record will be carefully filed for reference and will form a picture of your improvement. We need not do more than ask for your hearty coöperation in this experiment, you will be working for your own reputation, and what is more important, for the reputation of the school.

The material used consisted of extracts from Addison's "Essays," which were selected so as to be of approximately equal difficulty. The samples were as a rule slightly modified by omissions and insertions. This was necessary to avoid foreign terms, unfamiliar words, long and intricate quotations, etc. By these modifications a greater degree of uniformity in the weekly extracts was obtained. A short sample is shown, the originals consisted, in all cases, of not less than 350 words.

Sample. "The bridge, said he, is human life, look at it closely. Upon a more careful survey of it, I found that it consisted of seventy entire arches, with several broken arches which, added to those that were entire, made up the number about a hundred. As I was counting the arches the spirit told me that this bridge consisted at first of a thousand arches, but that a great flood swept away the rest and left the bridge in the ruined condition in which I now saw it. But tell me further, said he, what you see on it. I see crowds of people passing over it, said I, and a black cloud hanging on each end of it; as I looked more closely, I saw several of the passengers dropping through the bridge, into the great tide that flowed underneath it, and upon further examination, saw that there were many trap doors that lay concealed in the bridge."

The extract was printed and on a certain day of the week the test was administered to the whole school. Care was taken that the various sections that took the test early in the day did not communicate with the later sections. The following report from each individual taking part in the test was obtained, in addition to the actual extract that was typed in the five minute period: 1. Hours practised since last test,[1] (a) in

[1] The hours reported included the time spent in receiving instruction, as well as the actual time at the keyboard.

school, (b) out of school; 2. gross number of words typed in five minutes; 3. number of errors in words, punctuation and spacing; 4. net score on the basis of one point for each word attempted, less one point for each error. Before the test started very careful inquiries were made to find out whether the subjects had had previous practice in typewriting, if so, for how long, and what method of learning had been used. The scores as recorded in this paper are worked out from the data furnished by report 4, that is upon the total number of words attempted less one word for each error: in no word was more than a single error counted. The experimental errors and individual fluctuations are too great to justify the reduction of the scores to so fine a measure as number of strokes; the time that would be involved in such reduction would yield no adequate return.

At the beginning of this investigation, each pupil in the school who was studying typewriting, whatever his previous amount of practice either in school or out of school, was given the weekly test. This was continued until the group which was the most elementary had completed about 180 hours of practice. Meanwhile the more advanced sections had graduated. In this way the complete curve of learning was obtained for only one group; for the other groups merely a cross section of the curve resulted. The reason for the inclusion of all pupils was twofold, firstly it was hoped that the estimates of the previous amounts of practice of the sections would be sufficiently accurate to justify the compounding of results and comparison of groups. Secondly, if peculiar results were obtained with the most elementary section, during any time in the practice period, it was thought that useful corroborating evidence would be furnished by the other groups. In this present report the results recorded were obtained from the elementary section of pupils who were first given the test when they had practised for a period of 20 hours. They were consistently measured at two to three hour intervals, until they had completed a period of 180 hours practice. The data from the other sections have not been used in this paper, partly because there proved to be considerable doubt about the quantitative accuracy of the estimates of the periods of practice before the tests began, but largely owing to the fact that the results obtained from the group with 20 to 180 hours practice failed to reveal any peculiarity or anomaly which necessitated reference to the results of the other sections.

Reason for selecting 20 hours as the period of minimum practice

Dealing with the section that had just commenced typewriting, it was not feasible to measure improvement from the very early hours, for in the touch method, the attempt is first made to familiarize the learner with particular portions of the keyboard. Under these conditions the power of typing a continuous piece would be no test of improvement; in fact even after ten or fifteen hours there would be a decided tendency if the basis of scoring was one point off for each error to be zero or even negative. In the touch method, unless a certain point is passed, work of any reasonable kind is out of the question. This would not apply to learning by the sight method, where accuracy and steady improvement may be expected. In the touch system certain elementary habit groups have to be formed before the individual can attain a positive score at all; to this extent the scale is inadequate, for a zero score need not represent the zero of typewriting ability. However, any other method of scoring, unless we reduce to strokes correct, would suffer from greater disadvantages than this present method. It is perhaps legitimate to assume that an individual who on the average makes a mistake in each word, has, from the practical standpoint, something which approaches zero ability in the trait.

Final selection of subjects

About thirty-five individuals formed the group on which the first measurement was made after twenty hours of practice. Due to withdrawal, elimination, uncontrolled practice out of school, and lengthy absences only twenty [2] of these thirty-five individuals were consistently tested for the hundred and eighty hours of the practice period. If we consider the beginners in this section as a random sampling of the human material that commences the study of typewriting in commercial high schools, it is perhaps reasonable to suppose that the twenty individuals who completed the major part of the course form a slightly selected group. There are however certain counterbalancing factors. In most cases the elimination due to inefficiency, which is small, is not based primarily on lack of typewriting ability; in fact towards the end the withdrawals might well be those who, because of superior attainments in

[2] As a matter of fact nineteen, for the record of subject P ends at 110 hours; this record was included by mistake, but it has been retained in order to avoid recalculation.

stenography and shorthand, were commercially acceptable for early employment. The financial and social factors in the home—a large cause of the elimination—would be independent of typewriting ability. We may consider therefore, without making any appreciable error that the twenty pupils who completed the major part of the course were not selected upon any basis which would cause the results to lose any material value. The conclusions may be taken as typical of the normal rates of improvement for students of high school age and attainments, who are taught under the conditions prevailing in this school.

Tables and curves of improvement of the twenty subjects over period of practice from 20—180 hours

In order that these results may be used for more general purposes than the specific objects of this present study, it seems advisable to give in their complete form the material on which is based the curves of improvement for each individual. This is found in Table I. No record of errors is reported in this study, the two measurements which define any weekly test record are: 1. the total number of hours of practice since the last test taken; 2. the score on the basis of the number of words attempted in five minutes less the number of words in which errors were made, with one word subtracted for a spacing or punctuation error. If the reader is interested in the relationship between speed and accuracy a large number of typical results can be obtained from an earlier paper[3] which reported the positive acceleration in improvement which is found in many cases in the early part of the practice period. In this paper will be found a discussion of other studies which have investigated the learning curve in typewriting. For this reason any such discussion has been omitted from this paper.

[3] Chapman & Hills, *Journal of Experimental Psychology*, Vol. I, No. 6, pp. 494-507.

TABLE I

SCORE OF SUBJECTS (FIVE MINUTE PERIOD)

Hours Practised	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
20	38			34	27		46					30	41			59				
22	45	41		45	37						28				29					
24		52	55				55			28		30	53			61				
26	58			48	49	50		56	44		30	50	53	93	40		89	38	38	55
28		77	95		44	59		73	90	42	61	50		142		92	75	51	46	68
30	57	70	90	57	52		76				55	57			45	107				
32	76	88		62	59	64		87	88	51	82		48	103	54	106	84	58	59	88
34	92		101	64		72		86	105	70		58	64					66	74	97
36	144	82		106	62		58			82	80	73		154	66	123	92			
38		111	95		54	80	74	107	118		106		69		68		106	83	71	106
40	135			88						80		60	102			149				
42		86	131				78			109	82			180	82					
44	139			81	80	116	86	138	145			67	79		71	130	123	80	98	148
46	169	125	118	111	64			122	121	96	95	98		156						
48						85	62								72	144		91	67	
50	153	143		103	67					113	112	84	74	175	105		116			142
52	145		133	112		96	91	132	128		107		94			164		83	90	
54			150		80				151	142		126		162	81	150	136			
56	154	190		120	80	121	101	127		114	135	110	88					95	105	
58	181	205	128			115	92		154		150			186	115	197	128			226
60				122	89			173		156		127			162			130	94	
62	195	193	155	110	89	127	118		173		140	150		235		198	150	122	125	
64	185							226		145	135		116			186		146		
66			200	133	91	141	148		238			145		194	149		207		158	183
68	198	199		199	110		131			170	138	161				192		144		
70	197				157						147		134	186	138	198		143		
72				143		159	152	150		166	222	181					204		137	206

TABLE I—*Continued*

SCORE OF SUBJECTS (FIVE MINUTE PERIOD)

Hours Practised	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
74	205	195	197	166	125	....	142	179	....	....	160	186	....	231	153	177	....	153	....	....
76	....	191	....	....	....	136	158	....	180	167	....	202	134	....	131	191	188	166	147	241
78	229	....	....	186	145	....	....	200	182	169	185	....	....	....	....	191	185	....	....	209
80	267	....	....	177	122	151	158	....	....	....	180	207	....	201	158	....	....	159	135	....
82	....	215	221	....	....	148	168	201	212	....	223	192	123	209	....	216	203	195	....	225
84	193	....	....	....	156	....	....	....	....	....	....	....	....	....	145	205	....	....	134	....
86	....	204	....	....	....	157	190	176	223	182	....	....	....	207	....	219	200	222	174	227
88	....	....	217	167	....	....	202	....	228	206	214	193	124	225	193	....	....	180	....	235
90	230	215	....	174	137	131	230	220	....	....	214	....	....	....	165	....	206	....	193	....
92	223	....	....	....	....	....	....	....	236	235	....	197	....	230	....	211	....	....	....	....
94	....	209	214	....	153	160	....	238	....	207	192	201	135	....	....	....	216	....	186	221
96	....	218	....	....	....	....	212	....	....	....	239	....	....	....	200	218	....	....	....	....
98	234	230	....	185	148	176	....	....	....	....	....	205	158	....	....	....	239	183	158	....
100	250	....	225	168	164	....	233	....	199	202	214	222	....	201	....	210	....	180	....	....
102	....	....	....	....	....	161	200	....	....	....	....	....	....	....	198	246	206	....	....	281
104	256	....	....	187	185	....	....	254	....	202	215	229	156	....	....	242	....	....	....	....
106	....	230	198	....	....	....	261	....	226	....	....	....	....	....	234	....	....	....	222	....
108	....	....	....	....	169	172	274	....	....	222	205	....	....	220	224	....	....	199	....	....
110	248	....	....	196	....	....	....	233	275	....	....	....	160	274	....	....	....	209	196	....
112	....	....	213	....	....	....	240	....	248	206	....	....	....	....	215	....	233	....	....	242
114	206	....	....	187	191	....	....	244	....	....	207	207	....	297	....	....	....	199	209	....
116	....	261	....	....	....	209	....	228	247	....	....	....	187	....	....	....	232	....	....	284
118	218	....	193	179	217	....	235	....	....	226	198	....	....	254	208	....	....	....	....	....
120	....	....	....	....	....	180	216	217	....	....	....	205	....	....	....	....	231	209	197	....
122	....	....	....	....	208	....	....	....	206	....	226	....	....	251	197	....	....	....	....	244
124	250	....	215	206	....	184	....	221	....	194	....	189	164	....	....	....	272	187	181	....
126	....	231	....	....	....	....	227	237	196	....	....	....	....	....	....	....	....	....	....	232
128	259	....	....	226	177	180	....	....	....	180	228	....	....	229	182	....	....	178	....	....
130	....	....	240	....	....	....	252	....	226	....	....	234	....	....	....	....	230	....	206	....

TABLE I—*Concluded*

SCORE OF SUBJECTS (FIVE MINUTE PERIOD)

Hours Practised	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
132				260	204	159		238			263			208	220					261
134										211	263	230					213	199	209	
136	245	215	284	211		165	261	313	226				139		216					277
138								205		223		225		227				199	226	
140						189	287		298			201					243			
142			252				251		228	221				237	229			220		
144				238	245					226	215		142		233		253		236	272
146	284	238	231			180								249				224		
148	238			205	212						238		163						230	
150						186	283	247			231			261			321			
152	257			215		186						206					278			
154	246	247		207					245		236		176							
156			218		224					223					207					
158	269			249			256	304			266	228	166					232		235
160					220		298					219							233	
162	231		223	230	223	222	348		292		244			263	273					252
164	242	267		227			301	233		254	275	210	154				251			275
166	248	288	248	208	223	207			242		220	222						265		
168					262	202	286	245					177						280	
170	263			241		220		211	255	198	220			301				238		
172		280	248		238		293		262				180	237			264	238	218	
174			319			243				238					231		234			
176		258	256		248		268					204		269	233			232	248	
178					216	218			251				199	261			302			
180		264				206	276			206			204							
182					251												285	214	233	
184		280				205														
186			282			223											252			
188																				
190		292																		

From this table the curves of improvement of eight typical individuals have been drawn

SUBJECT B

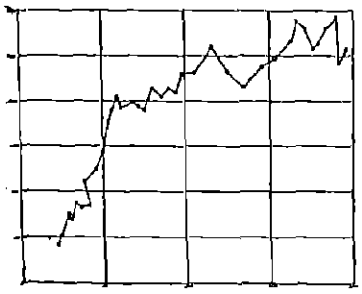


DIAGRAM 1

SUBJECT E

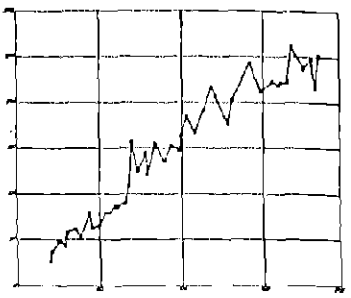


DIAGRAM 2

SUBJECT F



DIAGRAM 3

SUBJECT G

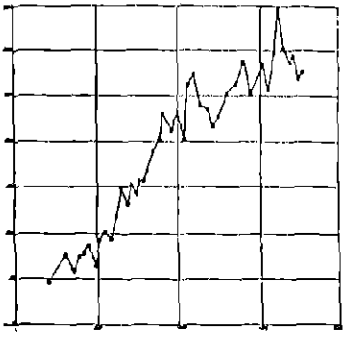


DIAGRAM 4

SUBJECT J



DIAGRAM 5

SUBJECT L



DIAGRAM 6

SUBJECT M



DIAGRAM 7

SUBJECT R



DIAGRAM 8

Care must be used before making generalisations on the basis of one or two points which form part of any particular curve. Each point on the curve is the result obtained from one test period of five minutes. The score of this period is subject to the fluctuation always found under such conditions, due to complex internal factors of interest, fatigue, interference, excitement, adaptation, etc. In addition any period which is represented by an erratic point may have been subject to the errors arising from 1. subject not stopping at signal given; 2. a copy which for some reason or another was much more difficult or much more straightforward for a particular subject than the average run of copy. This warning is given to prevent those who may work over these results from pressing any single values into the solution of problems, where the error of observation of the value may be too great. A combination of points however may be taken as a fairly reliable guide for such work.

An analysis of the shapes of the curves reveals among other things the following facts:

1. Individuals, reaching after the same period of practice the same degree of skill, do not by any means learn in the same manner;

2. there is no fundamental or typical curve to which all individual curves approximate;

3. short plateaus are present in the curve of many individuals, these plateaus are probably genuine, but they do not occur at fixed places which are the same for all individuals;

4. some individuals show distinct positive acceleration in improvement in the period from 20—60 hours;

5. some subjects are comparatively erratic in their variations from week to week, due to a complexity of internal factors, while others seem to vary little from the rate of performance which their own curve of improvement indicates is the true rate.

It is impossible within the compass of this paper to do more than mention these salient points, in a later paper the author hopes, if interesting facts result from the other material in his possession, to publish more evidence on some of these points.

Distribution table showing the relation between amounts of practice and scores at specified intervals of time

In the contingency table (table II), there are shown the scores made by the twenty subjects distributed according to the various practice periods. It should be noted that the twenty-five hour period includes twenty to twenty-nine hours of practice, while a score of twenty-five represents any value between twenty and twenty-nine.

A slight explanation is needed with reference to the cases included under any particular practice period; it does not follow that in each period the same individual will be included an equal number of times. Owing to practice out of school and to absences all of which were recorded, it may happen that whereas in the period 20-29 hours, A has five scores, during the period 30-39 A happened to have only four scores included. This factor however neutralizes itself and the averages obtained are almost identical with those that would result if within each period an equal number of scores for each individual had happened to be present.

Tables and curves showing scores at various intervals of practice

On the basis of distribution Table (II) the following scores for each ten hours of practice are shown in Table III.

(1) Average of the two lowest scores, (2) the 25 per cent point (counting from lowest individual), (3) median score, (4) the 75 per cent point, (5) average of the two highest scores. This tabulated material is exhibited graphically in diagram 9. In table (IV) the average score for each ten hours of practice is given, and the composite result shown in diagram 10.

The three salient features of the curve of improvement (diagram 10) are: (1) The initial linear relationship which extends over a period of 75 hours, (2) the abrupt turning point

TABLE II

WORDS WRITTEN CORRECTLY IN FIVE MINUTES

Hours Practised	25	35	45	55	65	75	85	95	105	115	125	135	145	155	165	175	185	195	205	215	225	235	245	255	265	275	285	295	305	315	325	355	345	355
25	4	8	12	14	3	4	1	3	...	...	...	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
35	...	...	2	13	9	9	12	5	10	2	1	...	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
45	...	...	...	...	5	4	10	6	2	3	4	4	4	1	1	...	1	...	...	...	1	...	...	...	...	...	...	...	...	...	...	...	...
55	...	...	...	...	1	1	6	6	5	8	7	4	4	7	2	1	2	2	1	...	1	2	...	...	...	...	...	...	...	...	...	...	...
65	...	...	...	...	...	...	2	2	...	4	5	5	9	5	2	3	3	7	2	...	1	1	1	...	...	...	...	...	...	...	...	...	...	...
75	...	...	...	...	...	...	...	...	...	...	1	6	6	7	6	2	9	7	6	...	2	1	...	...	1	...	...	...	...	...	...	...	...	...
85	...	...	...	...	...	...	...	...	...	...	3	2	2	6	2	3	3	6	11	6	8	1	...	...	...	1	...	...	...	...	...	...	...	...
95	...	...	...	...	...	...	...	...	...	...	...	3	1	3	1	2	4	3	6	8	3	10	...	2	2	2	1	1	...	...	...	...	...	...
105	...	...	...	...	...	...	...	...	...	...	...	...	...	1	4	1	3	4	6	3	7	4	2	2	1	2	1	1	...	...	...	...	...	...
115	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1	1	2	6	8	4	2	4	6	1	1	2	1	1	...	...	...	...	...	...
125	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1	2	8	4	5	3	6	4	2	2	...	1	...	...	...	...	...	...	...	...
135	...	...	...	...	...	...	...	...	...	...	...	1	...	...	2	...	...	2	5	5	4	7	2	2	4	1	1	...	...	1	...	...	...	...
145	...	...	...	...	...	...	...	...	...	...	...	...	1	...	1	...	2	...	2	2	9	6	3	3	...	1	2	1	...	...	1	...	...	...
155	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1	1	2	...	3	2	4	4	5	2	2	1	1	...	1	...	1	...	...	...
165	...	...	...	...	...	...	...	...	...	...	...	...	...	1	...	1	...	...	3	2	8	4	6	3	4	3	1	2	1	...	...	...	1	...
175	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1	2	1	4	2	9	5	4	6	...	1	1	2	1	...	...	...	...

at approximately 90 hours of practice, (3) the return to the second linear curve which extends from 90 hours to at least 180 hours of practice.

The main questions which arise from the abrupt turning point at approximately 90 hours are:

(1) Is this characteristic of the particular method of teaching (the Curry method)?

(2) Is it perhaps only true of the work of this one school?

(3) Is the school justified in spending the period from 90-180 hours for such small returns?

TABLE III

TABLE SHOWING SCORES AT VARIOUS INTERVALS DURING THE PRACTICE PERIOD

Hours practised	25	35	45	55	65	75	85	95	105	115	125	135	145	155	165	175
Average of two lowest scores	28															
First quantile	40															
Median	51															
Second quantile	69															
Average of two highest scores	118	149	175	216	237	236	251	239	278	291	263	299	313	313	325	329
No. of cases	50	65	48	58	52	54	54	44	41	40	38	37	34	30	42	39

DIAGRAM 9

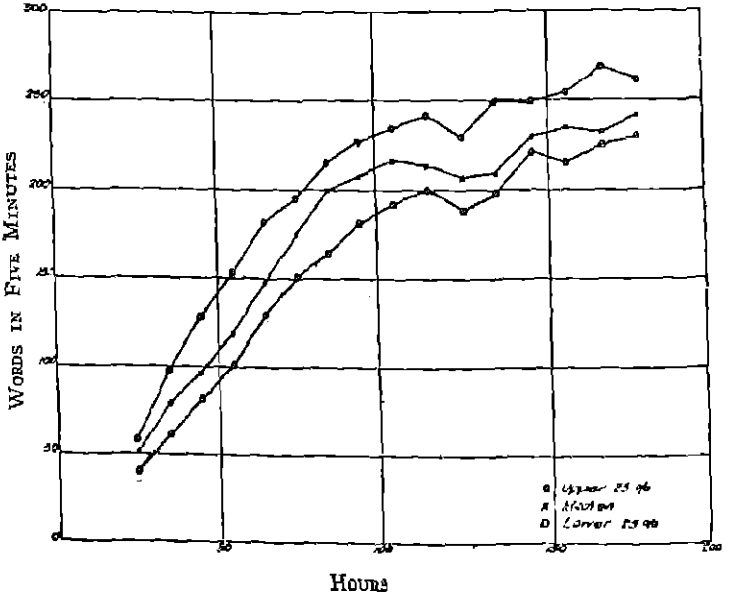
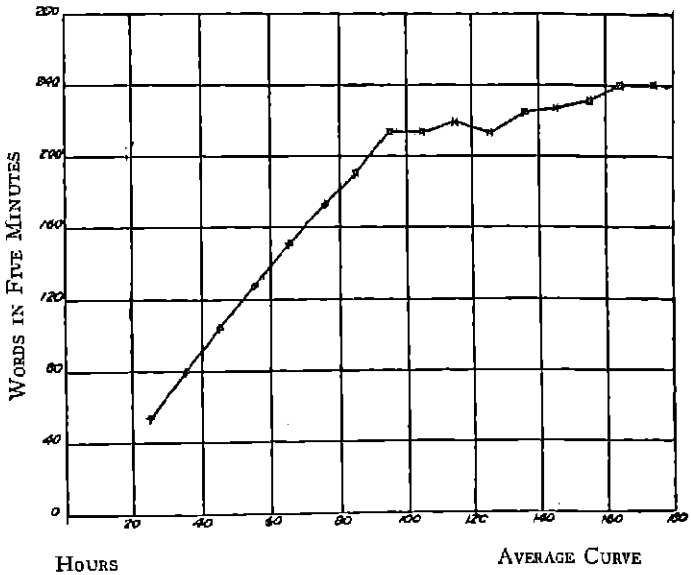


TABLE IV

AVERAGE SCORE AT VARIOUS INTERVALS OF PRACTICE

Hours practice..	25	35	45	55	65	75	85	95	105	115	125	135	145	155	165	175
Score, five minute period....	54	81	105	128	151	174	191	216	215	222	216	228	230	235	246	245

DIAGRAM 10



HOURS AVERAGE CURVE

Only one point need be made; as far as the writer can judge, there is nothing in the method used in this school, or change in type of instruction or aim of work, at ninety hours to suggest any such rapid alteration in the slope of learning curve. Most of the problems raised will only be settled by the publication of studies on similar lines which investigate improvement in typewriting, where other methods of instruction are employed.

While from the standpoint of diagnosis any equation of a curve with such an abrupt point of turning is worthless, for mere convenience the following equations are given, which have practical value for reference but no theoretical significance.

For period of practice, 20-90 hours $Y=2.3X$

For period of practice, 90-180 hours Y=180+4X

Where X is expressed in hours and Y is the average number of words typed in five minutes with one word deducted for each error. The incidental uses which can be made of the detailed results given in the tables are numerous, these have been left for reasons of space and because of the particular interest of the writer. However the examination of a fairly large group in a complex mental function over 160 hours of practice has yielded objective results which show how dangerous may be the attempts to construct a psychology of skill based on the short period examination of the rates of improvement of one or two subjects.

Forecasting of future skill in typewriting

In view of the work, which is being done in vocational guidance, which attempts to predict future ability in some specific occupation after training, it is interesting to analyze some of the results obtained in this paper. The question which is asked is: "To what extent is it possible from the initial success of an individual in typewriting to predict the degree of success after prolonged training?" To investigate this question the average score of each subject has been determined at three periods in the practice experiment. These periods are (a) 20-35 hours; (b) 76-91 hours; (c) 136-151 hours. The scores of each subject over each of these periods were averaged and they are shown together with the ranking in table V. The first period was chosen because it furnishes the first possible moment at which it was fair to make comparisons for reasons given earlier in the paper. The second period (76-91 hours) was selected because it falls before the point of change in the curve, which appears to take place at 90 hours. The third period was selected because it gives us some approximations to the final standing of the individual.

TABLE V

AVERAGE SCORES OF EACH INDIVIDUAL AT THREE POINTS IN THE PRACTICE PERIOD

Cross section of practice in hours	Average Score of Subject																		
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Q	R'	S	T
20–35.....	64	65	80	52	45	61	60	76	82	48	51	46	52	113	42	83	53	54	77
76–91.....	230	206	219	176	140	146	184	200	205	182	203	198	127	211	162	196	184	157	227
136–151....	250	227	256	218	228	180	271	255	251	223	228	213	148	244	226	272	244	231	274
Rank order 20–35 hrs.	8	7	4	13	18	9	10	6	3	16	15	17	13	1	19	2	12	11	5
Rank order 76–91 hrs.	1	5	3	14	18	17	11.5	8	6	13	7	9	19	4	15	10	11.5	16	2
Rank order 136–151 hrs.	4.5	12	4.5	15	10.5	18	3	6	7	14	10.5	17	19	8	13	2	16	9	1

The following points should be noted: 1. No test was made until twenty hours of practice in the function; 2. The test itself was an average of three to five measurements over a period of sixteen hours; 3. The test was in identically the same function as that in which later skill was desired.

The correlations between initial ability, intermediate ability, and final ability are shown below, using the method of rank square deviations:

Initial ability with final ability	.65
Initial ability with intermediate ability	.66
Intermediate ability with final ability	.66

These coefficients are reasonably high from the statistical point of view; that is a school, wishing to choose fifty pupils from a hundred who had practised 20 hours, might well consult the records at the end of this period of time. But if justice to the individual rather than a "statistical justice" to the group is necessary, it can readily be seen that the use of the initial ability as an indication of future skill would be very precarious. It is more likely that a mental test of fifteen minutes would give a better selection for the purpose of even the narrow function of typewriting and certainly for the whole work of the school.

The ordinary skill prediction test, when used for the purpose of vocational guidance, takes but a few minutes to administer, is often a matter of chance initial adaptation, and is usually in a very different function from that in which future skill is to be predicted. Tests of this type must be thoroughly tried out before their usefulness is admitted. For in contrast to the ordinary skill forecasting test, it must be remembered that the prediction test here described was a very lengthy one, after considerable adaptation, and in precisely the same function as that in which prediction was to be made. The low assurance value from the standpoint of individual guidance is revealed by the correlation coefficients. Vocational tests are not usually given for purposes of group selection but primarily for purposes of individual guidance. Therefore the distinction between "statistical justice" and "individual justice" becomes of great moment. Many who quote high correlation coefficients seem to forget this important consideration. To the man who is seriously wounded in action, the fact that he had a 99 per cent chance of safety is small comfort, but the same factor is a perfectly legitimate basis at headquarters for the calculation of replacements. The school must use every means in its power especially in the vital problem of vocational

guidance, to guard the rights of the individual while it adopts the necessary selections which are based on statistical averages.

Summary

The learning curve in typewriting has been studied over a period of practice extending from 20 to 180 hours. From a large amount of data, twenty full individual records have been selected for presentation. The following are some of the more salient results: 1. There is no fundamental typical curve of improvement; 2. prolonged plateaus are notably absent,[4] most of the subjects exhibit short plateaus which are probably genuine; 3. some subjects show positive acceleration in improvement in the period 20 to 60 hours; 4. subjects vary greatly in their erratic variations from week to week.

The norms of achievement at each ten hour period are given in six forms: a. the lowest point; b. the lower quartile point; c. median point; d average point; e. upper quartile point; the highest point.

The composite curve of improvement shows a most abrupt change of direction at 90 hours, from having an initial equation, (Y=2.3X) to a form, (Y=180+.4X).

The early results have been analyzed from the standpoint of their value in the prediction of future typewriting skill.

This study, which involved careful work each week for nearly two years, could never have been carried out without the continuous help of Mr. James S. Curry. Mr. Curry is the inventor of the method of instruction in typewriting which goes by his name. The subjects of this study were therefore under the supervision of the originator of the method. In expressing my obligation, I wish to join with the name of Mr. Curry that of Miss M. E. Hills, with whom this work was jointly started but whose absence from the city prevented its joint completion.

[4] Unless, as evidence seems to suggest, the period, from 95—180 hours, is considered as a long plateau, from which the subject, with great effort, can rise.

NATIONAL CONFERENCE ON AMERICANIZATION IN INDUSTRIES[1]

By Chester S. Carney

Introduction

About 200 persons assembled at the Atlantic Hotel, Nantasket Beach, Mass., June 23 and 24, to discuss the part which industry should take in the program of Americanization.

This meeting was an outgrowth of the Conference of Americanization Workers held in Washington in May under the auspices of the United States Department of the Interior. About three-quarters of the delegates represented industrial establishments either directly or through associations like the Associated Industries of Massachusetts and the National Manufacturers Association. The remainder came from such organizations as the Y. M. C. A., City and State Boards of Education, City Americanization Committees, and War Camp Community Service organizations.

This was a most happy combination of seers and doers, resulting in a series of inspirational addresses on what Americanization should accomplish, supplemented by suggestive discussions on the practical details of "how to do it."

What Is Americanization?

The keynote of the conference was the emphatic assertion that Americanization is a community activity, more organic than mechanical in its nature, that it means more than mere naturalization of foreigners. It is even bigger than an educational program of instruction in English, history, and civil government. It means the assimilation of our foreign born population into all phases of our national life, involving a change in the attitude on the part of the native American, as well as a development of the ideals of the foreigner. "You cannot become an American overnight, by an initiation process, the way you become a Mason or an Elk." "It is not like

[1] This report is reproduced here by kind permission of Scovell, Wellington & Company, Boston, New York, Chicago, Cleveland and Springfield, Mass., Industrial Engineers and Scientific Management with whom the author is associated.

changing a suit of clothes." "We must be more concerned with the quality of citizenship rather than with the quantity."

It was repeatedly pointed out that each racial group has its contribution to make toward our mutual welfare; that we have much to learn from the "stranger within our gates;" that no line can be drawn between Americans and immigrants: we are *all immigrants* originally; and that we must never allow ourselves to indulge in the "I-am-holier-than-thou" attitude of raising the foreigner up to our level, but rather that we have a great destiny to work out which we cannot accomplish without his help. We want to work *with* him, *not* for him, realizing that he is an asset, not a liability.

Should Attendance at English Classes Be Compulsory?

The problem of how to retain class attendance drew the fire from all the batteries. Representatives of several large plants argued at some length for compulsion. The analogy of compulsory attendance at public school was drawn. It was asserted that some workers did not know what was good for them, and that those who would not attend English classes were undesirable employees. "The voluntary attendance plan had been tried, and had failed. Attendance started at 200 and dropped to 20."

Opposing the benevolent paternalists were the champions of democracy, proclaiming the essential sanity and good intentions of the majority of mankind; that it was rare, when the matter had been properly presented to the men, that anyone refused to attend class; and that the few unwilling individuals would finally yield to the pressure of public opinion without necessitating the use of the big stick. This argument prevailed to such an extent that a vote of 100 to 8 was taken in favor of voluntary attendance. That this was not wholly indicative of the sentiment of the entire conference was evident in the open forum Tuesday afternoon, when the compulsionists were reinforced by late arrivals and the democrats were weakened by the loss of their leading orator. But a vote was not taken, and the majority still seemed opposed to compulsion. During the discussion a very able mediator distinguished clearly between the men who are successful because of their powerful personality, in spite of the fact that the theories they hold are false, and men whose theories may be fine but who fail to "get them across."

Should Employees Be Paid for Time Spent in Class Attendance?

Few arguments were raised on either side of the question. Most seemed to agree that when classes were held during working hours the hourly rate should be paid, and when classes were held after working hours no payment should be made. A census of the delegates showed 18 plants holding classes during working hours, 36 plants holding classes after working hours, and 6 dividing the class period between employee's and employer's time. Three plants actually charged an enrollment fee, which in two cases was returnable at the end of the course.

What Relation Should Exist Between the Teachers of English to Non-English Speaking Workers and the Public School Authorities?

One of the two most discussed issues raised at the convention was the extent to which the work of teaching English to non-English speaking workers should be turned over to the public school authorities.

On the one side were those who had in mind the general principle that teaching was the function of the public schools, and production the function of industry; that therefore all educational work should be headed up and supervised by the public school authorities.

On the opposite side were those who had found particular instances where the public schools had not lived up to their civic obligation, so that the work would have been neglected had not philanthropic or industrial organizations come to the rescue.

Between these two camps there was no essential disagreement. It was practically the unanimous opinion that cooperation with the public schools was desirable. All admitted that in some communities the public school had not been ready to assume the responsibility, and sometimes had failed to accomplish the work it had undertaken. But the faults, it was argued, were faults of a locality, and not of a principle. As one opportunist put it: "We all know that 'the cathedrals are built of the stones of the country,' that conditions in one locality make one plan a success, while the same plan would fail utterly under the conditions which obtain in another locality. But we must agree that as a general rule, supervision by the public school is desirable."

Practical Hints On How To Do It

In the discussion of practical details some valuable suggestions were brought out.

From several sources came descriptions of graduation exercises which help the new citizen to realize the significance of his enfranchisement; that his "Second Papers" are in a different class from his vaccination certificate, with which they were frequently confused by applicants for war work in a certain plant! In one place the graduates were addressed by the Mayor, the Governor, and a representative of the Federal Government, after which they were personally introduced to these officials.

In another a pageant was held, where the "Goddess" of the city welcomed each racial group into her family expressing her appreciation of the gifts which each group contributed, and maidens in folk costumes responded, laying at her feet symbols of these gifts.

The necessity of reaching the foreigner in his own language was often stated. One of the most practical papers, by a man who has organized Americanization work in several plants, presented the following program:

1. *General propaganda*—through foreign-language papers, pay-envelope stuffers, bulletins and pamphlets describing the plan in detail printed in the native languages of the foreigners.

2. *Group mass meetings,* addressed by interpreters who are prominent in the national group.

3. *Personal interview* by a representative of the management, a high grade man who will explain that the aim is to help the worker to find his way through the difficulties of red tape attached to the process of naturalization.

4. *Training for citizenship.* The laboratory, or "direct" method was advocated; the first seven of ten lessons to be devoted to what the government does for the individual, the last three on what the individual does for the government. The teacher must know the difficulties which confront the applicant for citizenship, especially the court procedure. One teacher conducts a mock-court as a final lesson of the course, showing just what is to be expected.

5. Arrange for *filing papers* in court. The foreman is a better witness than the saloon keeper.

6. Welcome to citizenship should be a social event in which plant and civic officials should participate.

In teaching English, four grades are useful:

I. Illiterates in their native language.

II. Those who cannot speak English, but can read and write another language.

III. Those who speak English, but can read and write it very little.

IV. Those who can read and write English. These are ready for citizenship course.

One speaker pointed out that instruction in English was the first step in safety work.

Another stated that classes in English were held for convalescents in the hospital where the work helped to fill their idle time.

In one plant where the company was conducting classes twice a week during working hours the students of their own accord requested a third session to be held on their own time.

A large rubber company held "Recognition" exercises Sunday afternoon, which 1,200 people attended, and at which 400 students received diplomas in recognition of 75 or 150 hours' attendance at classes.

In addition to the class work, all recreation activities and the work of the visiting nurse were declared to be valuable points of contact by which Americanization can be forwarded.

There was repeated emphasis upon the importance of the foreman's active cooperation in all efforts for Americanization, particularly in following up absentees.

One plant has accomplished fine work in cooperation with the International Institute of the Y. W. C. A. Rooms were rented in the district where the foreigners lived, and classes were conducted in English and domestic science for the wives and children of the workers.

In another plant dances are held frequently, with folk dances taught by foreigners and American dances by the native born.

An interesting suggestion was that of modifying the 8% alien income tax to apply to all who have not taken out first papers, and 4% to all who are not citizens, the proceeds to be used for Americanization work.

A representative of the State Board of Education outlined a plan by which instructors were to be chosen from within the works, and trained to teach English to adult immigrants in a class conducted by the board of education. He pointed out the advantage of such teachers to the plant in promoting

esprit de corps, and asserted that teaching English to adult foreigners was a special problem in itself.

One speaker asserted that Bolshevism was a product created in New York by our failure properly to Americanize our immigrants.

Addresses On Other Subjects

In addition to the discussions of Americanization, several worth-while addresses were made by men who have been doing conspicuous work during the war. Mr. Howard Coonley, president of the Walworth Company and formerly vice-president of the Emergency Fleet Corporation, told of their experiences in handling disagreements between management and men, which were settled in conferences between seven representatives of the men and seven representatives of the management.

He expects great volume of business this fall, with accompanying shortage of labor.

Mr. W. Graydon Stetson spoke on the program for thrift and saving, which will probably be continued for at least five years more, he believes.

Mr. B. Preston Clark, of the Plymouth Cordage Co., spoke most interestingly of the two great movements resulting from the war: the increasing power of the people, and the increasing value of leadership. He quoted the Arabian proverb,—"The camel-driver has his thoughts, and so has the camel, too," and asserted that we need more than ever before to consider the thoughts of the "camel." The futility of theorizing without intimate sympathetic knowledge of the conditions he illustrated by quoting:

> "The toad beneath the harrow knows
> Exactly where each tooth-print goes,
> The butterfly flitting by the road
> Preaches contentment to the toad."

He is hopeful of an ultimate solution of the Americanization problem, because this conference has shown very clearly that we possess the two essential qualities—imagination and humanity.

Mr. Frederic C. Hood, president of the Associated Industries of Massachusetts, in his opening address declared that management has two branches, one the management of materials—which is applied science—and the other the management of men—which is applied religion. He called attention to the need of a higher grade of leadership, and asked whether managers were prepared to supply this leadership, or must the workers look elsewhere. He asserted that 75% of the cases which came before the War Labor Board came from

industries suffering under absentee management. He emphasized the strategic importance of the teacher in the Americanization program, and advocated the payment of better salaries. He urged all to read:

"*Back to the Republic.*" Harry F. Atwood (Laird & Lee).

"*Money and Prices.*" J. Lawrence Laughlin (Chas. Scribners Sons).

An able paper written by Mr. Alvan McCauley, president of the Packard Motor Company (who could not be present), pointed out the need for greater coordination and cooperation among the various bureaus of government and the volunteer organizations concerned with Americanization. He declared that "there is a confusion of tongues" and much working at cross purposes.

APPENDIX A

Resolutions Adopted by the Conference

"We, the representatives of the educational forces of industry, recognize that industry has a definite part with the other forces of the community in initiating and organizing Americanization work; therefore,

"*Resolved,* That instruction in English for non-English speaking people should be carried on in cooperation with the public educational forces, provided such forces are prepared and will assume responsibility. We pledge our aid in our respective communities to bring about this cooperation.

"*Resolved,* That non-English speaking employees attending English classes in industry should attend such classes voluntarily and without compensation.

"*Resolved,* That every industry employing non-English speaking people formulate a definite policy regarding Americanization work, and that such work can best be done when a responsible person is charged with its direction.

"*Resolved,* That we, as a representative group of industries, unanimously disapprove making naturalization a condition of employment, and recommend that every community establish at least one school for citizenship."

APPENDIX B

List of Pamphlets on Americanization

A number of pamphlets were exhibited, which would be suggestive to any plant planning to install classes in English or citizenship. Such a plant will receive help from the organizations referred to below.

City History Club of New York, 105 West 40th Street, N. Y., have a program using the laboratory method.

Associated Industries of Massachusetts, 1034 Kimball Bldg., Boston Mass. "Americanization in Industry—How to Organize English and Citizenship Classes," by George F. Quimby. 31 pp.

Americanization Bureau, Pennsylvania Council of National Defense. George W. Pepper, chairman, Philadelphia. "Americanization in Pennsylvania." 40 pp.

Americanization Committee, City of Duluth, Minnesota. "How to Become an American." 24 pp. 15 cents. (Printed also in Italian.)

Y. M. C. A., Lawrence, Mass. "The Citizen: First Lessons in United States History for Immigrant Men." 31 pp. 25 cents.

Charles H. Paull, Harvard University, Cambridge, Mass. "Americanization: A discussion of present conditions, with recommendations for the teaching of Non-Americans." (A report submitted to the Solvay Process Company.) 27 pp.

Federal Board of Vocational Education. "Advanced Course in English for Foreign-born Men." 31 pp. 1919.

Raymond F. Crist, Deputy Commissioner of Naturalization, Washington, D. C. "Students' Textbook: Standard Course of Instruction for the Preparation of the Candidate for the Responsibilities of Citizenship." 132 pp. 1918.

Department of University Extension, Massachusetts State Board of Education. "Federal-State Program for Immigrant Education." 20 pp. January, 1919. "English for American Citizenship." 12 pp. July, 1918.

Old Colony Trust Company, 17 Court Street, Boston, Mass. "Getting Together." Extracts from addresses on Bolshevism, by A. J. Peters, N. B. Endicott, and W. H. Taft, May 20, 1919. 18 pp.

Cleveland Board of Education. "Lessons in American Citizenship." Raymond Moley. 65 pp. 1918.

Edwin A. Rumbull, General Secretary Civic Education Association, Buffalo, N. Y. "New Citizens' Handbook." 34 pp. (Printed in Polish, Italian, Hungarian, etc.)

International Committee of Y. M. C. A. has prepared extensive course in English for foreigners.

The General Electric Company has prepared moving-picture film on Americanization, which may be borrowed by those interested, on application to Mr. M. R. Griffeth, 84 State Street, Boston.

MINOR STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

I. A COMPARISON OF COLORED AND WHITE CHILDREN BY MEANS OF A GROUP SCALE OF INTELLIGENCE.

By S. L. Pressey and G. F. Teter.

I. *Problem and Materials.* Colored children show a greater school retardation, less acceleration, and average older for a given grade, than do white children.[1] There is some evidence to show that, grade for grade, they do poorer school work.[2] Negro children give ratings, when tested by the Binet and Point Scales, averaging below white children.[3] Measurements of special abilities have shown the colored children to do relatively well in tests of the more simple processes (as cancellation, rote memory) and most poorly with tests of the more complex abilities[4] (as opposites, analogies, sentence completion). There is some evidence that colored children have more active imaginations and more ready associations of an uncontrolled type than do white children.

There would thus appear to be the beginnings of an exact science of racial differences, as shown by school children. But—as has been well pointed out recently[5]—the proper evalu-

[1] See for instance, Perring, L. F. A Study of Comparative Retardation, *Psychological Clinic*, Vol. 9, 1915, pp. 87-93; Mayo, M. J. The Mental Capacity of the American Negro, *Archives of Psychology*, No. 28, 1913.

[2] See for example, Mayo, as above.

[3] Strong, A. C., Three Hundred Fifty White and Colored Children Measured by the Binet Simon Measuring Scale of Intelligence, *Ped. Sem.* Vol. 20, 1913, pp. 485-515; Phillips, B. A., The Binet Tests Applied to Colored Children, chiefly Kaffirs, *Tr. Sch. Bull.*, Vol. 12, 1915, pp. 122-3; Redfern, L. H. and Rich, S. G., Binet Tests on South African Natives—Zulus, *Ped. Sem.* Vol. 24, 1917, pp. 373-83; Sunne, Dagny, A Comparative Study of White and Negro Children, *J. of Appl. Psychol.*, Vol. 1, 1917, pp. 71-83.

[4] See especially Pyle, W. H., The Mind of the Negro Child, *School and Society*, Vol. 1, 1915, pp. 357-60; Ferguson, G. C. Jr., The Psychology of the Negro; An Experimental Study, *Archives of Psychology*, No. 36, 1916, pp. 138; and Sunne, Strong, Redfern above; also Woodworth, R. S., Comparative Psychology of Races, *Psychol. Bulletin*, Vol. 13, 1916, pp. 388-96; and Thorndike, E. L., *Educational Psychology*, Vol. III, pp. 206-24, 1914.

[5] Some Suggestions Relative to a Study of the Mental Attitude of the Negro, *Ped. Sem.* Vol. 23, 1916, pp. 199-204.

ation of such data is not easy. The unsatisfactory school work of colored children may be, in part at least, merely an indication of lack of adaptation of the schools to their special needs. It may surely be questioned whether tests given by white examiners to colored pupils can give reliable data for a comparison of the races. There may even be some doubt as to whether, with examiners of their own race, the reaction of the colored children to the test situation would be quite the reaction of white children. In any case, the fact must not be forgotten that the colored children come from a social and psychological environment which is subtly but powerfully different from the environment which molds the mentalities of the white children; standards of living, moral and religious ideas, sentiments and interests and ideals, are all pervasively different. It should be said at once that in the present instance the tests were given by white examiners, to colored children in schools showing no attempt at adaptation to the particular needs of the race, and in communities where there is sufficient race consciousness for the colored people to be a distinct class by themselves. But certain of the differences found seem too marked to be explainable as the result of such factors alone. There is, further, a sufficient variety of material available with regard to the subjects studied to make possible some little analysis of the situation.

The material consists of measurements, with a group scale of intelligence recently developed at Indiana University,[6] from all the children from the third through the twelfth grades in two schools for colored children in two small Indiana cities—a total of 187 cases in all—together with a study of the age-grade placement of these children. The examining was done by the senior writer and his wife. The scale used may be briefly described as consisting of ten tests, each of twenty items, each test being applicable roughly from the third grade through high school or from 8 to 16 years in age. The individual tests will be described briefly, later, in connection with an analysis by test.

The paper will take up in turn the following questions: (1) Grade for grade, how do the colored children compare with the white children? Do the colored schools have their own standards as to the amount of innate ability necessary for work in a given grade? In studying this question the colored schools will be compared with the grade distributions from one of these two cities (City A)—a total of 1,022 cases.

[6] Pressey, S. L. and L. W., A Group Point Scale for Measuring General Intelligence, With First Results from 1,100 School Children, *Jr. of Appl. Psychol.* Vol. II, 1918, pp. 250-269.

(2) Age for age, how do the colored children compare with white children? The scores made by the negro children will be compared with results from a total of 2,800 white children tested in these two cities and one other city (Cities A, B, and C). (3) Do the colored children show a distinctive make-up of abilities? In studying this question the results on the separate tests of the scale will be compared with norms for each test obtained from the white children of City A.

II. *Results.* The composition of the grades in the colored schools may be briefly indicated. The number of children tested in each grade was as follows:

Grade	3	4	5	6	7	8	9	10	11	12	Total
Number	31	36	49	28	14	5	7	4	12	1	187

The figures suggest a banking up of cases in the first three of these grades, and a high retardation. The colored schools showed, in fact, a total retardation 90% greater than the retardation for the white children of City A, and an acceleration less than half as great. The colored children also averaged ten months older per grade. If we turn now to the scores made by these children on the tests, we find that only 36% of the colored children scored at or above the median for their grade in the white schools, 13% scored above the 75 percentile, 51% below the 25 percentile. To put it another way, the colored children of a given grade average about at the average for the grade below theirs in the white schools. Apparently the colored schools require less intelligence, for admission to a given grade, than do the white schools. Grade for grade, the children in the colored schools average ten months older than the white children. Nevertheless, the grade averages are distinctly lower in the colored schools.

But the fact that the colored schools distribute their children differently in the grades, as regards age, makes it impossible to obtain, from the grade averages, any satisfactory evidence as to the amount of difference between colored and white children in mental endowment. For such a purpose children of the same age should be compared. The number of cases tested at each age was as follows:

Age	7	8	9	10	11	12	13	14	15	16	17	Total
Number	1	11	20	24	25	27	23	21	14	12	9	187

After careful study of the age-grade distributions it was decided that the most reliable ages for comparison were ten, eleven, twelve, thirteen, and fourteen. Below ten there were too many cases who were in the first and second grades (and so not tested) to give a fair sampling; after fourteen the duller children drop out of school. Of 120 colored children

within these ages, only 14% score at or above the median for the white children of the corresponding age. Only 3% of the colored children score above the 75 percentiles of the white distribution. But 54% of the colored children score below the white 25 percentile. To put it another way, the colored children average at about the average for white children two years younger.

When compared age for age, the colored children thus make a poorer showing than when compared grade for grade. Apparently the colored schools, having to deal with poorer mental material, have tried to keep up grade standards by retarding their children more—but have only partially succeeded.

The scores of the colored children were next analyzed by test (using again, data from the five ages above mentioned). The per cents scoring at or above the medians for the white children, on each of the ten tests, are as follows:

Test	1	2	3	4	5	6	7	8	9	10
Per cent	33	30	29	15	29	22	14	13	22	16

The colored children do relatively well on the first, second, third, and fifth tests. The first is a test of rote memory; the children are read such lists as "chalk, man, bag, book, sister," and then asked to write the word that came after "chalk." The fifth is a test of logical memory; the children are read a newspaper passage and are then asked questions about it. The second is a test of practical judgment; the test consists of such lists as "CHAIR;—arm, cushion, legs, rockers, seat," and the children are told to "underline, in each list, the two things which the thing named in big letters at the beginning of the list is *never* without." The third is a test of practical arithmetic; the children are read such problems as "If you had 25 cents and bought two erasers at 3 cents each and a bottle of ink at 5 cents, how much would you have left?"

The tests on which the colored children make the poorest showing are four of the five "literary" tests of the scale. Two of them involve controlled association, the fourth (Opposites) and the tenth (Analogies); one of them requires a knowledge of moral terms and one calls for verbal ingenuity (number 8, Disarranged Sentences). The sixth test (Word Completion—the test consists of such words as "a—it—me—ic" and the children are expected to fill in the missing letters) and the ninth test (Practical Information—the test contains such lists as "bill, coin, check, draft, stamps," and the children are asked to "put a line under the safest way to send money,

and a cross after the least safe way") give results, with the colored children, midway between these extremes.[7]

III. *Discussion.* The above results are undoubtedly influenced by some of the prejudicial factors mentioned at the beginning of the paper. The examination was given by white examiners; the tests were framed for use with white children. But certain of these differences between white children, and colored children living in the same community and taught under the same school system, are too marked to be wholly thus explained. A poor average ability seems unmistakably indicated—a poor ability inadequately compensated for in the schools by a greater retardation, and so leading to lower grade standards as well.

It is natural, however, to seek an explanation for the lower scores in some special defect rather than in a general incapacity; each race might be supposed to have both weak and strong points. Analysis by test gives little support for such a theory; the colored children show poorer ability than the whites on every test. But there are distinct differences from test to test; twice as many colored children make scores above the medians for the whites on the first, second, third, and fifth tests as on the fourth, seventh, and eighth. The colored children do best in a test of rote memory. They do poorest in a test of verbal ingenuity and a test involving knowledge of abstract terms. It is easy to generalize, and conclude that colored children do best on tests pertaining to very concrete problems, or making demands upon the more simple mental processes. But such generalizations help us very little in getting at the real facts as to mental differences. It is more pertinent to note, in this connection, that the first test has been found poorest of the ten in differentiation of feebleminded and subnormal cases, the eighth most differential, and that the fifth and first are least differential of superior ability, the seventh and tenth the most differential. There is the definite suggestion here of a more elementary and less highly developed ability among colored children.

But one keeps coming back to the notion that the important racial difference may be, after all, emotional and tempera-

[7] It should also be noted that the per cent passing each test above the white medians shows a fairly regular decline from the first to the last test; that is, there is a suggestion of a fatigue curve. One cannot help but wonder whether there is not evidence here of a greater fatigability on the part of the colored children. Fatigability is surely not the only factor at work; the last tests are also tests of the more complex processes. But this dropping off in scores toward the end of the examination is distinctly interesting.

mental. Possible evidence of a greater fatigability on the part of the colored children has already been mentioned. The present study offers little material relative to such differences. But the writer is convinced that measures of such differences can be obtained. He feels strongly that the possibilities of group "tests" as instruments for research have hardly yet been realized. Group tests for the measurement of persistence, of effort and interest, are, he believes, a likely development of the near future.

IV. *Summary.* Measurements, obtained from the children in two colored schools with a group scale of intelligence, show that:

1. The colored children of a given grade average at about the average for white children one grade below theirs.

2. The colored children of a given age average at about the average for white children (in the same city) two years younger.

3. Analysis by test shows the colored children to average below white children of the same age on all the tests. The colored children rate best on a test of rote memory, poorest on a test involving differentiation of abstract terms and a test of verbal ingenuity. It is suggested that in general colored children may do relatively well in dealing with concrete and routine problems, and poorly where abstraction or mental reconstruction is called for.

4. Study of racial differences by means of group tests of persistence, and of interest, is suggested.

II. A STUDY OF COUNTRY CHILDREN IN (1) A GOOD AND (2) A POOR FARMING DISTRICT BY MEANS OF A GROUP SCALE OF INTELLIGENCE

By S. L. Pressey and J. B. Thomas.

I. *Materials and Problem.* The materials of the study consist of results from 270 country school children tested by one of the writers (Dr. Pressey) and his wife in a poor rural district in southern Indiana, and 268 country school children in a fairly good agricultural district in the central part of the state, examined by a graduate student in Psychology (Mr. Thomas). The first group was examined in the spring of 1918, the second in the fall of that year. The scale employed in the preceding study was used here; and comparisons are with the norms referred to there.

The children examined were almost entirely of American stock. Special difficulties qualifying the findings, such as language handicap, were thus not encountered. It should be said at once, however, that the data from the two country districts includes, besides children in the one room country schools, children in the schools of several small villages, trading centers for the country round about. In the first district there is also a small quarry town of perhaps 200 inhabitants, and in the second district a slightly larger mining town. Both quarry and mine obtain their labor largely from the country near, so the community centering around them can hardly be considered essentially different from the surrounding country. In fact, in this mingling of agriculture and small scale industry, the districts may be said to be distinctly typical of southern Indiana.

In working over the data three questions have been considered. (1) Age for age, how do country children compare with city children? In studying this question the results from the two country districts have been combined and comparison made with results from the 2,800 city children, already mentioned. (2) Do the country children show distinctive results on the separate tests? The comparison here is with norms for each test obtained from City A. (3) Is there any evidence to show that the less intelligent people live in the poorer country district? Comparison has here been made of each

of the two country districts with the norms from three cities previously referred to.

II. *Results.* The total number of country children at each age ran as follows:

Age	8	9	10	11	12	13	14	15	16	Total
No.	17	59	86	81	78	76	49	39	18	503

Because of the factors of grade selection mentioned in the preceding study, it was considered wise to use only ages 10-14 inclusive in these comparisons. It soon appeared, however, that elimination in the country began earlier than in the city. This affected particularly the group examined in March, the older boys having already left school to help in the spring plowing. Only country children from ages 10-13 have therefore been used in the comparisons. The per cents, at each of these ages, testing at or above the median for city children of the same age are as follows:

Age	10	11	12	13
Per cent	29%	33%	21%	25%

Of the total of 321 country children, of these four ages, 27% test above the age medians of the city children; 8% score above the 75 percentile; 42% rate below the 25 percentiles.

Again analysis by test is naturally looked to for aid in understanding these findings. In making this analysis the results from the same four ages have again been lumped, and the per cent of the country children scoring at or above the median for their age, on each test, has been found.

Test	1	2	3	4	5	6	7	8	9	10
Per cent	38	35	39	26	40	29	25	33	33	31

The most important feature seems to be the lack of such marked differences from test to test as were found in the similar analysis of the colored data. The colored children do almost three times as well on the first test as on the eighth. The poorest results shown by the country children are on test 7; the best showing is 40% on test 5.[1] The country children do not thus present a distinctive make-up of examination, but show a slightly poorer ability on all tests.

It is not infrequently asserted that in the country districts a constant selective process is going on, the poorer, less intelligent stock being pushed back more and more into the hill country where the land is poorest, while the more able, assertive elements of the population obtain the best land and the best opportunities. That is, on this assumption, there

[1] For a description of the separate tests see the previous study.

should, in an agricultural community, be a positive correlation between land values and intelligence. The two districts from which the data of the present study were obtained seem not ill suited to test this hypothesis. District 1 is hilly, the soil is poor; markets are hard to get at. The land in District 2 is much better and markets are much nearer. It has been estimated roughly that the land in the second district is worth three times as much as that in the first.

A comparison of the school children in the two districts by means of a mental test is thus distinctly interesting. For the four ages used above, 20% of the children in District 1 rate above the medians for the city children as compared with 36% in District 2—the better district; 6% in the poorer district score above the 75 percentile for the city children, as compared with 11% in the better district; 48% in the poorer district rate below the 25 percentile for the city children, as compared with 28% in the better district.

These figures are probably not as reliable as one would wish. The schools in the poorer district were examined in March, which is close to the end of the year for the six months' schools. Already the children who could be of most assistance in the spring work on the farms had left school, in many instances. Because of this the figures from District 1 are possibly a little too low.[1] As has already been said, the two districts were not examined by the same examiner. The examinations in District 2 were made by the least skilled examiner, and it is not improbable that the children in this district profited by his inexperience and his desire to be wholly fair to his children. But it is hard to entirely explain away the differences between these two districts by such special circumstances. At least, the results are of suggestive value,—are of decided interest in indicating ways in which results from a mental survey may be of use in the study of sociological problems.

III. *Discussion.* A word remains to be said as to possible explanations of the lower scores made by the country children. The obvious conclusion would be that the country children were less intelligent. The writer feels that such an explanation is decidedly not wholly adequate. For one thing, country children do not meet the test situation in the easy matter-of-fact way usual with city children. They are distinctly shy with strangers, especially if they feel that the strangers come from the city; they adapt much less readily to the special

[1] The country teachers usually say, however, that it is the duller children who thus leave school before the end of the term.

procedures and types of problem involved in the usual measurements of intelligence. That is, they show distinctly less "testability," though not necessarily less intelligence in meeting the problems of their own environment.

Further, the tests used were all verbal; they were presented on paper and involved the constant use of a pencil. The children were not as dextrous with their pencils as city children. In many instances, there were only two or three pencils in a school; the written work was on the blackboard or on slates. Awkwardness, some embarrassment, and slow writing were the consequences. Moreover, verbal tests, such as were used, can hardly be expected to give a total measure of ability in a community where the work is almost wholly manual. Performance tests and, perhaps, a different mode of approach would seem necessary. It was urged in the preceding paper that the usual forms of intelligence tests are inadequate for the study of racial differences; it is urged again that these same tests need not a little supplementation in studying groups from very different environments.

IV. *Summary.* 1. A group of unselected country children rate about a year and a half in mental age below city children.

2. Analysis by test shows nothing that could be considered distinctive.

3. The children in a good farming district average above children in a poor farming district.

4. It is urged that the usual type of intelligence tests does not give adequate measures of the ability of country children; performance tests and materials more relevant to their environment are needed.

BOOK REVIEWS

H. Addington Bruce. *Nerve Control and How to Gain It.* Funk & Wagnalls, New York and London, 1918. p. IX and 307.

The purpose of this book, as stated by the author in the preface, is to aid primarily those persons who are suffering from nervous invalidism or feel themselves approaching it. Hope is also expressed that it will prove useful to normal individuals who wish to gain a precise knowledge of the fundamentals of personal hygiene.

With simple non-technical language Dr. Bruce presents in a clear wholesome manner the facts of mental hygiene that have been derived from the many medical, physiological and psychological researches carried on by the leading investigators in all parts of the world. Assuming that nervousness is not a disease in itself but rather a symptom of ill health which may have resulted from any one of many physical or mental causes, he proceeds to discuss these causes one by one in the form of little friendly talks. The topics include the effects of such bad habits as worry, fear, anger, selfishness, trouble seeking, etc., and how to overcome them; the value and proper use of foods, the beneficial effects of the right kind of exercise, the importance of assuming a correct posture, the vital relation of fresh air and sunshine to health, the benefits of rest and relaxation and the necessity of a right outlook, besides many others.

The book closes with a plea for the establishment of free institutions by the state, municipality or philanthropic individuals, for the care of persons suffering from functional nervous disease who have not the financial means to take treatment at private sanitoriums. In the author's words "it is a disgrace, it is socially criminal to leave the nervous poor to shift for themselves."

This contribution comes from one who, from long training and experience in the field of mental hygiene, is able to speak with authority and it can, without question, be recommended as a sound and practical guide to the art of right living and thinking.

I. A. Field.

Roy Franklin Richardson. *The Psychology and Pedagogy of Anger.* Warwick & York, Baltimore, 1918. p. 100.

In this little book, Professor Richardson has summarized the results obtained from a dozen Clark University students, trained in introspection, who attempted to make full observations of their emotions for a period of three months. Their reports are classified under the following chapter headings: Mental Situations Stimulating Anger; Behavior of Consciousness; Disappearance of Anger; Conscious After-effects; Educational Function.

Prof. Richardson's subjects find that while the psychic background of anger may be a general irritation tending to focus on some particular person or object, or the inducement of negative self-feeling for which the anger compensates, the balance of evidence goes to prove that Wundt's conscious fore-period is ordinarily absent, the emotion

of anger seizing the person suddenly without any awareness of its genesis. The conscious response to the emotion may be one of three types:

1. "Attributive reactions," in which the angry person indulges in all types of mental imagery whereby he expresses his feelings (in imagination only), or builds plans for future revenge, etc.

2. "Contrary reactions," in which religious and moral beliefs control his conduct so that he restrains his wrath and "loves his enemy."

3. "Indifferent reactions," in which the helplessness of the angry person leads him to make the best of the situation and assume a philosophical attitude of indifference.

By far the largest percentage of anger reactions fall under the first class. The second class conforms so closely to the psycho-analytic mechanisms of repression and defense that it is a pity the author did not take time to point out the analogy and give us his opinion concerning the application of the Freudian principles developed in connection with the sexual emotion to other instincts and emotions.

An affective attitude of pleasantness or unpleasantness may be concomitant with anger, accordingly as it does or does not find adequate relief in expression. Only in cases where it is fully and completely vented upon its object or in some substitutive activity does anger finally disappear. Controlled anger tends to return to consciousness repeatedly. Here again it is to be suggested that the introspective results confirm expectations which would be created by an application of the psycho-analytic rubrics.

The after-effects of anger are varied,—pity, shame, anxiety, contrition, pleasure, increased friendliness, relief, curiosity, fear, etc., are listed, and seem to depend on the complex of factors which go to make up the stimulus situation and the type of subjective reaction. This ambi-valent swing into another emotion has been explained by Cannon and other investigators on a physiological basis, and the chapter would be more comprehensive had Prof. Richardson chosen to supplement his introspective material with a reference to these findings.

The chapter on educational functions is rich in pedagogical suggestions. It is impossible to over-emphasize the motivating power of held-over anger attitudes in human conduct, and the teacher who knows how to arouse and guide anger aright is indeed wise. Each individual pupil must be studied, for where anger may prove an incentive to greater effort in one, it may be the source of discontent and a surly, mischievous attitude in another. If it can be roused against the proper obstacles, and sublimated into arduous labor, it becomes a vital factor in education.

PHYLLIS BLANCHARD.

NOTES.

September 1st, 1917, the Community Service Society, Youngstown, Ohio, organized and opened The Children's Service Bureau as one of its departments with a separate staff and board of directors. The plan and work of the Bureau was outlined and promoted and is now being directed by Mr. J. M. Hanson, Secretary of the Community Service Society. It has been privately supported by Mr. H. H. Stambaugh, who has just died, and who while living, quietly but generously and effectively used his personal influence and finances in developing constructive social programs.

The Bureau was established to be the general clearing house for work in behalf of children in Mahoning County. It was founded on the belief that by early study of the "problem children" in the schools and by intelligent action in meeting their individual needs, most of them can be saved to themselves and to society. This, therefore is what the Bureau has undertaken to prove to the community. It has worked in close co-operation with, and has supplemented, the Public Schools and the Court of Domestic Relations, and intends to continue doing so until provision is made from tax funds for a scientifically directed educational program. One school of nearly 1,000 pupils representing 20 nationalities, was chosen as a demonstration center. Here every child entering the lower grades is studied to determine his physical condition, his mental abilities and defects, and the environmental influences. These studies are now in progress and when completed will serve as a basis for recommending to the school such medical and social corrective measures, such re-classification of children and such an altered educational procedure as this investigation indicates and warrants. Careful follow-up work in the homes is done by the school nurse and school extension worker.

The Paidological department, under the direction of Dr. H. H. Young, has been making intensive studies of children referred to it by parents, the public schools, the Court of Domestic Relations, the Children's Home, Civilian Relief of Red Cross and the various charitable and social agencies of the city. Dr. Young has been officially appointed by the Board of Education as their examining psychologist and, although not officially appointed, has also been doing the psychological work for the Court of Domestic Relations.

The Bureau, because privately supported, is able to serve these two important public institutions and to help work out a few of the many problem cases they have in common. As a result of this correlation the rapidly accumulating evidence shows how the work of the Juvenile Court and of the schools is one and the same, and that there is no such a real differentiation in function as the existence of these as two separate institutions indicates and emphasizes. Time and again, the child referred by the schools is found to have a court record, and the one referred by the court is found to be the source of much annoyance and a problem in the schools. With such cases, the best social adjustment can be made if the child's past record and present condition be investigated, evaluated and interpreted by some one agency which serves as a clearing house in co-ordinating the various isolated attempts of independent agencies.

When in 1916 the Institute for Public Service began issuing from New York City its weekly post card bulletin *Public Service* and other educational reviews it was stated that its financial backer could not be named. For the first time the chief donor's name is now published,—Julius H. Barnes now U. S. Wheat Director and since 1917 president of the U. S. Grain Corporation which stabilized wheat prices during the war.

Mr. Barnes' interest in "learning by doing" made him consider starting a boys' school. He finally decided to promote "learning by doing" everywhere instead of in one spot and started the Institute for Public Service to try out circulating facts about best practices in school work. The offices are at 51 Chambers street, New York city from which center it has conducted field surveys, clipped educational and lay journals for best practices; compiled and distributed such pamphlets as High Spots in New York Schools, Rainbow Promises of Progress in Education, War Fact Tests for Every American, Teacher Personality Charts, Graduates Who Win and many others; answered questions from teachers and superintendents.

Mr. Barnes himself as U. S. Wheat Director will value the co-operation of the nation's teachers in securing the further patriotic help of children and their parents in circulating and using the truth about our bread supply. His offices are at 42 Broadway, New York city from which inquiries will be promptly answered.

It was Mr. Barnes' suggestion, too, that during this next year the Institute issue from time to time teachable facts about bolshevism. Several bulletins were issued last year and a great many letters answered. For copy of these bulletins and further information address William H. Allen, director, 51 Chambers street, New York city.

Dr. W. W. Charters, Dean of the School of Education of the University of Illinois, will on September first become Professor of Education and Director of the Research Bureau for Retail Training at the Carnegie Institute of Technology. He will also serve as consultant on curriculum problems arising within the various technical and vocational divisions of the Institute. Dr. J. B. Miner, who has been Acting Director of this Research Bureau, will continue with the Bureau and will also teach some advanced courses in vocational psychology.

Dr. Kate Gordon has been promoted to an Associate Professorship in the Division of Applied Psychology at the Carnegie Institute of Technology, and has been given leave of absence for a portion of next year to continue the investigations for the California State Board of Charities begun during the first semester of last year.

Major C. S. Yoakum, Associate Professor of Applied Psychology at the Carnegie Institute of Technology, has become Director of the Bureau of Personnel Research, formerly called the Bureau of Salesmanship Research. Colonel Walter Dill Scott continues as Associate Director of this Bureau, part-time, but he will be resident in Evanston during the greater portion of the year.

Dr. Harry D. Kitson, Instructor in Psychology at The University of Chicago, has accepted the position at Indiana University, made vacant by Professor E. C. Lindley who accepted the presidency of the University of Idaho.

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. IV DECEMBER, 1919 No. 4

MENTAL RATING OF JUVENILE DEPENDENTS AND DELINQUENTS IN ALABAMA.

By W. D. PARTLOW, M. D., Superintendent of Alabama Insane Hospitals, *Tuscaloosa, Alabama*; and

THOMAS H. HAINES, M. D., Professor of Medicine (Nervous and Mental Diseases), College of Medicine, Ohio State University, Columbus, Ohio.

This paper constitutes a report upon the value of group intelligence rating tests, and the relations of the same to individual examinations made by the Binet-Simon method as a part of a mental survey of the four industrial schools of Alabama. This opportunity came to us by request of the Chairman [1] of the newly constituted Board of Control and Economy. This Board recognizes it to be a matter of fundamental economy for it to know the kinds of mental equipment the boys and girls of these industrial schools possess, in order that the real work of the institutions may be forwarded in the best possible manner.

The schools investigated are four in number,—the Alabama Boys' Industrial School, at East Lake, with a population of about 307 white boys, the State Training School for Girls and the Mercy Home Industrial School for Girls, both at Birmingham, with about 45 and 30 white girls respectively, and the Alabama Reform School for Juvenile Negro Law-Breakers, at Mt. Meigs, with 264 colored boys. These children range in age from seven years to twenty.

[1] The request for this work on the part of the State Board of Control and Economy was met by the Alabama State Board of Health and the National Committee for Mental Hygiene. These agencies financed the survey.

Backwardness of Industrial School Population

It is a well known fact that boys and girls selected by Juvenile Courts, and other social agencies, for residence at such institutions as these industrial schools, are backward in school work,—that they are in large measure what we call dullards and feeble-minded.

TABLE I

Distribution by age and School Grade at the Alabama Boys' Industrial School

School Grades	Ages in Years													Totals
	7	8	9	10	11	12	13	14	15	16	17	18	19	
Primer	1	1	2		1						1			6
I		4	2	2	2	2	4	2			2	1	1	22
II			3	2	7	6	6	7	4	4	9	1	2	51
III			1	4	6	6	7	14	8	4	2	2		54
IV		1			4	3	15	12	6	7	3			51
V				1	1	3	5	13	12	10	1	2		48
VI							3	10	9	7	9	1		39
VII								1	8	3	3	1		16
VIII							1	2	6	2	2			13
IX									1	1	2			4
Totals	1	6	8	9	21	20	41	61	54	38	34	8	3	304

Table I exhibits these facts from the point of view of school classification for the Alabama Boys' Industrial School. 304 of the 307 boys examined, for whom school classification was stated, are classed according to age and school grade. If it is fair to assume that a child in his seventh year should be in the primary grade, in his eighth year in the first grade, in his ninth year in the second grade, and in his fourteenth year in the ninth grade, we find that there are only eight of these

304 children who are up to, or beyond, the expected grade. They are shown below the diagonal line in the table. Only 30 other children of the 304 are in the grades immediately below these, or one year retarded, according to these expectations. The great mass of children point in the direction of the nine second-grade seventeen year-olds.

The modal age is fourteen years, and the modal grade is the third. Likewise the maximal number of third graders is fourteen at fourteen years of age. The modal grade for the fourteen year-olds is the third, and the modal age in the second grade is seventeen years with nine already alluded to. Likewise the median age of the second grade is fourteen years.

The group of this Industrial School population is, in fact, so loaded with retarded children that this condition cannot be attributed to the accident of poor school facilities. Especially true is this when we consider what large numbers of these boys come from the cities where excellent school facilities are provided.

Group Intelligence Ratings

Experience with group tests in the applications of psychology to the organization of military forces had convinced us of the labor-saving feature of this device in such a survey as this. The military tests were not available when this work was planned. Considering the large amount of illiteracy amongst these boys and girls, as indicated by the large numbers classed in the lower school grades in Table I for the white boys, it seemed advisable to prepare two forms of group examination, one for children of the fourth school grade degree of literacy in English, and the other for children of the third school grade and below. The form prepared for those able to read English, and designated 'Form A' of Group Intelligence Rating Tests, is constructed very much on the lines of the Alpha Examination used in the Army. It is somewhat shortened and introduces one performance test, namely, the mazes. It consists of eight parts, (1) Directions Test, (2) Arithmetical Performance, (3) Mazes, (4) Same-Opposite, (5) True-False, (Disarranged Sentences), (6) Analogies, (7) Choice of Reasons for Stated Facts, and (8) Arithmetical Series. The total possible score for these eight tests is, according to the method adopted, 193, and the actual working

time of the examinee about twenty minutes. It is, therefore, an easy matter to go through the whole performance with a group, including the filling in of the headings of the paper, in the space of an ordinary school period of forty-five minutes.

The other form of Group Intelligence Rating, Form X, designed for children who have not much reading knowledge of English, consists of five parts, the first four of which involve no reading of English whatever. The first is a Directions Test, the second is a game at store keeping, marking goods in letters which stand for figures (a letter digit test), the third is a maze test, and the fourth consists of twenty unfinished pictures. A cloth blackboard is carried to aid in the explanation of the second, third, and fourth tests. The first maze is on the blackboard entire, and a child is asked to volunteer to show the group how to get through this maze, by coming up and taking the chalk and making a mark from the arrow on the left to that on the right of the maze. The first unfinished picture, a girl without a mouth, is on the blackboard as on the examination book, and a volunteer is asked to come up and finish this picture. The transition from unfinished pictures to unfinished sentences (Test 5) is made as easy as possible by the fact that both are unfinished. The unfinished sentence is by all odds the most difficult of these five tests for these relatively unlettered children, because it involves reading of English. The total possible score in this Form X of Group Intelligence Rating is 182. The actual working time of the examinee is about seventeen minutes.

In order to be as economical of our time as possible, and at the same time disturb the routine of the institutions as little as possible by our work, we proceeded first to make group examinations of the intelligence of the boys and girls of these schools by using these two forms.

The same two forms were used about the same time on an aggregate of 1012 children in the schools of Jackson, Mississippi, ranging in age from five to thirteen, and in school grade from the primer grade to the seventh, or highest grammar school grade. In the same way 765 colored children in the schools of Jackson, ranging in age from five to fourteen, and in school grade from the primer to the seventh, were rated as to intelligence by these two forms. The percentile distribution of the scores made by these Jackson school children constitute our measuring instrument for the performances of the boys and girls, white and colored, of the Alabama Industrial Schools.

GROUP INTELLIGENCE RATINGS OF CHILDREN IN ALABAMA INDUSTRIAL SCHOOLS

TABLE II

DISTRIBUTION OF SCORES IN A & X FORMS OF GROUP INTELLIGENCE RATING BY AGE OF CHILD AND BY PERCENTILE MAXIMA OF PUBLIC SCHOOL CHILDREN. ALABAMA BOYS' INDUSTRIAL SCHOOL

Per-centile	Ages in Years													Totals
	7	8	9	10	11	12	13	14	15	16	17	18	19	
100					1		1							2
90							1	1	3	2	4	1		12
80				1				2	4	4				11
70							1		4	2	2			9
60							2	2	2	2	2			10
50					1	1	2	3	3	1	2			13
40			1		1	1	2		4		2			11
30				1		1	1		1	2	2			8
20		1	1		2	3	2	10	8	7	2			36
10			1	2	4	5	2	17	4	5	2	2		44
5				2	5	2	4	8	6	3	4	1		35
3		4	3	1	1	4	2	2	1	3	1			22
2	1	1	2	2	6	4	22	16	15	7	11	4	3	94
Totals	1	6	8	9	21	21	42	61	55	38	34	8	3	307

Table II presents the percentile distribution of the performances of 307 boys of the Alabama Boys' Industrial School rated by the percentile distributions of performances of Jackson school children. Since we had very small numbers of children beyond nine years of age in the third grade in the public schools, nine years is the highest for which we have reliable percentiles in the X form of examination. Likewise our figures for the A form of examination comprise children from nine to thirteen years, inclusive.

219 of these 307 boys are beyond thirteen years of age. More than forty per cent of these boys took the X form of

examination by virtue of their retardation in school work. There is necessary, therefore, a certain amount of projection of expectation of performance with advancing age. We have to anticipate what a thirteen year-old might reasonably be expected to do in the X form of examination from what seven, eight, and nine year-olds have done, and likewise we have to anticipate what a sixteen year-old might be expected to do in both the A and X form from what younger children have done.

The distribution of the scores of the Boys' Industrial School inmates have been made in Table II according to this simple plan. Any score falling below the lowest two percentile performance of Jackson school children is placed in the two percentile group under the age of the child making it. Any score falling above the two percentile, and below the three percentile limit, is placed on that level, and so on, above three and below five is placed in the five level, and above five and below ten is placed in the ten level. By this simple plan of stratification we have put before us another striking exhibition of the retardation of the minds of these boys. 151 of these boys put up mental performances below the limits of the lowest five per cent of public school children of the same ages and race. That forty-nine per cent of these boys are so incapable mentally is really a more striking exhibition of the problems confronted by the management of such an institution than are the actual numbers of feeble-minded which we present later.

This classification of intelligence of the boys abundantly justifies the school grading as set forth in Table I.

Of course we have constantly to bear in mind the fact that the group examination does not detect the lazy individual who will not put forth his best effort under the conditions of such examinations. However, the conditions of the test were exactly the same in the public schools and in these industrial schools.

In the same way Table III exhibits the distribution of scores of the 75 girls of the two Training Schools according to the percentile stratification of performances of public school children.

33 of the 75 girls fall below the five percentile level of public school children, and only one of the whole number makes a score above the median performances of public school children, whereas 44 of the 307 boys exhibit scores above the median performances of public school children. This seems to indicate in general a lower grade mentality amongst the white girls of these reform schools than amongst the white boys.

TABLE III

DISTRIBUTION OF SCORES IN X FORM OF GROUP INTELLIGENCE RATING BY AGE OF CHILD AND BY PERCENTILE MAXIMA OF SCHOOL CHILDREN. STATE TRAINING SCHOOL FOR GIRLS AND MERCY HOME INDUSTRIAL SCHOOL FOR GIRLS

Per-centiles	Ages in Years														Totals
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
100															
90															
80															
70															
60							1								1
50						2	1	1	2		3	1			10
40						1			1	2					4
30							2	1	1	1	1	1			7
20							1	2	4	4	1	2			14
10							1		2	2	1				6
5						2	1	1	2	1	1				8
3															0
2	1		1		1		3	2	6	5		2	3	1	25
Totals	1		1		1	5	10	7	18	15	7	6	3	1	75

Table IV shows the stratification of the scores of 264 negro boys in the Alabama Reform School for Juvenile Negro Law Breakers. This stratification is made by comparison with the performances of 765 negro children in the public schools of Jackson, Mississippi. It is a very different distribution from what would be presented had they been compared with the performances of white children. 221 of these 264 boys, or 84%, score below the median performances of children of their same ages and race. Judged by the standards of their own race, 84% of these boys are below the average of intelligence. 16% therefore, are above average intelligence for their race, whereas 14% of the white boys, and 1% of the white

TABLE IV

DISTRIBUTION OF SCORES IN A AND X FORMS OF GROUP INTELLIGENCE RATING BY AGE OF CHILD AND BY PERCENTILE MAXIMA OF PUBLIC SCHOOL CHILDREN. ALABAMA REFORM SCHOOL FOR JUVENILE NEGRO LAW BREAKERS

Percentiles	Ages in Years												Totals
	8	9	10	11	12	13	14	15	16	17	18	19	
100				2			1		1				4
90				1			1	1	1				4
80			3			2	4		1				10
70					2			5	3	1			11
60			2	1	1	3	2	1	2	1	1		14
50		1				5	4	2	1				13
40				5	2	3	5	4	2	1			22
30		1		1	8	15		3	2	1			31
20	1			1	7	16	11	15	14	4		1	70
10			2		3	3	15	10	12	10	2	1	58
5			1	2	3	1	7	5	5	1	1	1	27
Totals	1	2	8	13	26	48	50	46	44	19	4	3	264

girls are above the average intelligence for their ages and race.

85 of the colored boys, or 32% of the total number score below the limits of the lowest 10% of public school children of their race, and 27, or 10% below the limits of the lowest 5%.

The scores of the colored boys do not bunch so closely in the lowest five percentile strata as do the white boys. A much larger number is found between the ten and twenty percentile limits than below the five percentile limit.

But it is evident that all three of these groups,—white boys, white girls, and colored boys, are over-loaded with dull children, white boys 86% below average, white girls 99% below average, and colored boys 84% below average.

The distribution, however, within this lower half of the field of mental ability is very different for the three groups.

Below the ten percentile level we find white boys 64%, white girls 52%, and colored boys 32%.

Below the five percentile level we find white boys 49%, white girls 44%, and colored boys 10%.

On being compared with school children of the same race, the colored boys are, therefore, from all three of these angles, better endowed than the white children. While they have not significantly larger numbers, or percentages, above the average intelligence of their race, they do rank higher in the low groups, or strata. Whatever is the selection process of the courts sending these boys and girls to these schools, they evidently do not select negro children so closely as they do white children along the lines of racial dullness, or lack of intelligence.

Individual Examinations

Throughout the experience of the personnel of the Division of Psychology in the Surgeon General's office in separating the feeble-minded from the competent in the draft army, it was found necessary to follow the scoring of group performances by individual intelligence examinations in order to make satisfactory diagnoses and to accumulate the evidence necessary for discharge from the service on account of mental inferiority. For the most part the feeble-minded were found in the so-called 'E Group' of scores resulting from group intelligence rating. This group constituted roughly the lowest five per cent of the draft.

As to percentages found Terman states that 16,266 men out of 1,059,767 examined between May 1, 1918 and September 1, 1918, had a mental age of eight years, or less.[2] This constitutes one and six-tenths per cent of the gross number in the drafts under consideration.

Yerkes[3] states later that of 1,726,000 men examined prior to January 1, 1919, 45,563 were of a mental age of less than ten years. This is 2.6% of the drafts under consideration. As, however, only 27,052 of these were recommended for (1) discharge as feeble-minded, (2) labor battalions, or (3) development battalions, this is probably the truer measure of mental deficiency as found in this draft. This number, 27,052, is 1.6% of the drafts under consideration. It seems likely, therefore, that something like one and one-half per

[2] L. M. Terman, *Psychological Bulletin* XV, June, 1918, pp. 177-187 "The Use of Intelligence Tests in the Army."

[3] R. M. Yerkes, *Science* March 7, 1919, p. 226, "The Measure of Utilization of Brain Power in the Army."

cent of the men who passed local exemption boards and were conducted to camps as reasonably suitable material for making fighting men were so defective in mentality that there was no hope of developing them into serviceable factors in the military organization.

The defects of the plan of rating adults by mental age, and judging of their efficiency thereby, are exhibited by these figures. It is evident, despite the striking presentation it enables one to make to the lay mind of the defect of a mentally inferior person, that it does not represent the facts in the case to say of a given individual twenty-five years old that he has the mind of a child of seven, or of nine years.

The different ages represented in the developmental period by these children in the Alabama Industrial Schools complicates our problem considerably beyond that confronting the military psychologist. We have to treat each age group up to sixteen as a distinct group and to compare the score of each with the scores of normal children of the same age and race.

As a matter of practical procedure, being compelled to proceed with our individual examination work before the stratification of scores of public school children was completed, we arranged the scores in each age group, for each race and sex, and for each form of the examination, in order of their magnitude, and proceeded with the children making the lowest scores up in the lists until we felt satisfied we had reached a point beyond the level of feeble-minded performance. This empirical procedure probably involved a little more labor than would have been necessary had we had at our disposition the stratification of public school children's performances. However, with this classification of public school children's performances in hand we would have proceeded in exactly the same way, only being enabled to stop a little short of the point to which we mounted in each group. For mental age rating we used the Stanford Revision of Binet-Simon Tests.

Value of Group Tests

The first point in favor of the use of group intelligence rating methods in such institution surveys as these herein reported, consists in the facilities afforded for a quick and fairly accurate intelligence rating of every individual in the survey. It is a matter of decided value to the Superintendent and management of an institution, just as to the personnel officer in a division of the Army, to know of each individual that he is above, or below the average in intelligence, and how much above or below he is. It is of service to such management

to know that one-half of the children are as low down in ability to adapt themselves to their environment and to learn how to do things as the lowest five per cent of public school children of their same ages and race.

There is no question about the labor saving effected by this device when one is setting about the survey of an institution population with the specific object in mind of sorting out the feeble-minded. It is a much more effective procedure to make group tests and get the scores than to discuss intelligence capacity and possibility of mental deficiency with the Superintendent and teachers. This situation has not changed in the twenty years since Binet called attention to the very poor ability of the average teacher to diagnose, or to aid in the diagnosis, of mental deficiency in the case of any child in her room. Any who have ever tried this method of discovering which children in a given group should be submitted to an individual examination for mental age rating and social prognosis know how exceedingly unsatisfactory it is. Such persons know that one must practically examine every child in order to be reasonably certain he has not passed over some mental defectives.

By this method we found it unnecessary to extend our individual examinations beyond about one-third of the white boys and white girls, and a little over one-half of the colored boys. This is some curtailment of labor.

Of the boys and girls actually examined individually, approximately one half were found to be feeble-minded.

As stratification of intelligence of school children proceeds to larger numbers, and we can state with more definiteness the expected performance of the lowest two per cent, the lowest three per cent, and the lowest five per cent of children of each age through the developmental period, we shall be able to come closer to the feeble-minded group by the group method of intelligence assessment. As this approximation is made, more labor in individual examination will be saved.

Of course this is an asymtotic approximation, the special factor precluding diagnosis of mental defect by the group method alone being the impossibility, already referred to, of detecting that species of volitional interference with the production of good results in an individual, otherwise capable, under the conditions of the group method of intelligence testing. There will always be individuals who will do less than they are capable of performing. Some such will, in any group, have to be interviewed and examined individually in order to demonstrate that they are not as poorly equipped as their group intelligence rating results indicate.

TABLE V

RELATION OF GROUP SCORES TO DIAGNOSES RESULTING FROM INDIVIDUAL EXAMINATIONS

TOTAL NUMBERS OF

(1) Scores as of lowest five percentile public school children.
(2) Children assigned mental age by Stanford Binet-Simon Test.
(3) Diagnosed as Feeble-minded (institution cases).
(4) Diagnosed as Border-line (probably institution cases).
(5) Diagnosed as not Feeble-minded as result of individual examination and mental age rating.

	White Boys	White Girls	Colored Boys
Numbers scoring in Group Rating as of Lowest 5 Percentile	151 (49%)	33 (44%)	27 (10%)
Numbers Assigned Mental Age by Stanford Binet-Simon Examintion	94 (31%)	25 (33%)	140 (53%)
Numbers diagnosed as Feeble-minded	44 (14%)	16 (21%)	57 (22%)
Numbers diagnosed as of Border-line Intelligence	13 (4%)	3 (4%)	42 (16%)
Numbers diagnosed as not Feeble-minded by Individual Examination.	37 (12%)	6 (8%)	41 (16%)

Table V exhibits for the three classes of children examined the total numbers rating themselves below the five percentile limit of public school children of the same race, and, in parentheses beside these figures the percentages which these numbers constitute of the total numbers of those examined. The next line presents the numbers and percentages of each group submitted to individual examination; the next, the numbers and percentages of each group diagnosed as feeble-minded; the next, the numbers and percentages diagnosed as of border-line intelligence; and the last, the numbers and percentages submitted to individual examination but pronounced not feeble-minded. It should be stated parenthetically at this point that considerable numbers of boys and girls

submitted to individual examination and found not to be feeble-minded were found to be constitutionally inferior persons, or epileptics, or psychopaths, and for these reasons likely to prove constant public charges on account of criminal tendencies growing out of their abnormal characters. Many children were seen individually and partial mental examinations made, for whom no mental age assignments were made.

An out-standing feature of this Table is the very large percentage of negro boys given individual examinations. It will be noticed also that a considerably larger percentage of these negro boys are diagnosed as feeble-minded than of the white boys, and a very much larger percentage of the negro boys are diagnosed as of border-line intelligence than of either white boys or white girls. It is a fact that we are without data in regard to the reasonably expected performances by a Binet-Simon form of examination of negro children at the various periods of development. The standardization of individual intelligence rating has been done with white children. We do not know what to expect of negro children. As a consequence of this defect in our methods of applied psychology we found ourselves more at a loss in making prognoses of these negro boys in Alabama than in respect to the white boys and girls.

The group tests with Jackson school children, and the work with group tests in the Army, have shown unquestionably the backwardness and inferiority of the mental development in the negro as compared with that in our white population, but we are entirely without a guide as to the measurement of this inferiority in terms of years and months. We have no standards by which to judge what degree of intelligence is required of a negro boy of a given age in order to warrant us in predicting his ability to be such that he will be able to manage himself and his affairs with prudence in this civilization of ours. Consequently, we have taken refuge in classing a larger number of them as of border-line intelligence. At the same time we have found it necessary to examine individually a far larger number of them, and have classed a larger percentage of them as feeble-minded than of either white boys or white girls.

Certain it is that if we judge these negro boys by the standards established for white children, we should find two thirds of the population of the Alabama Reform School for Juvenile Negro Law Breakers subjects for state care as mental defectives, and we should also find very large percentages of negro children in the public schools subjects for state care.

Manifestly the practical results of allowing negroes of such inferior intelligence, judged by white standards, to undertake to manage for themselves, are not disastrous. We are not likely to have to look after larger percentages of negroes as state wards than we do of the white population. It is, therefore, a scientific desideratum of the highest import that we should have a standardization of the Binet-Simon Tests to the development of negro intelligence.

Another feature of contrast between the races manifest in our results is a wider percentile distribution of the negroes as distributed according to the strata of performance of public school children. Of the white boys and girls given individual examinations, only three had scores in the ten percentile group, that is, above the five percentile limit, two of them boys, eleven and twelve years old, made, on individual examination, intelligence quotients of 100 and 104; the other, a girl seventeen years old, made an X score of 102, but on individual examination made an intelligence quotient of 59, her mental age being 9.4. This girl, it should be said, was called in for individual examination more on the strength of a statement by the Superintendent than on account of her showing in the group tests, and she was rated as a psychopathic personality really suffering from an incipient psychosis. As a result of this condition she shows two different mentalities, one in the group test which is good, and another in the individual examination which is inferior.

On the other hand, the negro boys called for individual examination and found to be feeble-minded range high in the percentile stratification of their group scores.

Table VI exhibits these facts, showing that the majority of the five percentile group (63%) were diagnosed feeble-minded, only 43% of the ten percentile group, 20% of the twenty percentile group, and 3% of the thirty percentile group. Likewise, decreasingly small proportions of these groups were called for individual examination.

The distribution of examinees and the resulting diagnoses of these negro boys shown in Table VI indicate that the standard of feeble-mindedness for the negro boy, which the examiners had in mind in their work, cuts through the percentile distribution of group scores of negro children at about the twenty percentile level. This twenty percentile level in negro children seems to correspond closely to the two percentile level in white children. 28% of the five percentile white boys are found feeble-minded, and none above the five percentile group. 45% of the five percentile white girls are found feeble-

minded, and, as just explained, no real case is found above this group. On the other hand 36% of the lowest twenty percentile group of the negro boys are feeble-minded, and one case is found above the twenty percentile level.

TABLE VI

DISTRIBUTION OF GROUP SCORES OF NEGRO BOYS GIVEN INDIVIDUAL EXAMINATIONS, AND NUMBERS OF EACH GROUP FOUND FEEBLE-MINDED, BORDER-LINE AND NOT FEEBLE-MINDED

Percentile Group	Numbers in Groups (Table IV)	Numbers Examined Individually	Found Feeble-minded		Diagnosed Border-line	Diagnosed Not Feeble-minded
			Numbers	Per cents of whole groups		
5 Percentile	27	27	17	63%	8	2
10 Percentile	58	51	25	43%	14	12
20 Percentile	70	47	14	20%	17	16
30 Percentile	31	13	1	3%	2	10
40 Percentile	22	1			1	
80 Percentile	10	1				1

REASONS FOR THIS SCATTERING OF THE NEGRO FEEBLE-MINDED

We cannot satisfactorily explain this disparity of results of correlation between group and individual examinations for the white and colored populations. The following facts, however, may be taken in partial explanation of this disparity:

(1) Smaller numbers of negro children were examined in the public schools; 765 stratified scores as contrasted with 1012 for white children. It was also noted in this stratification that there were numerous cases of maxima for older children falling below maxima for younger children at the same levels.

(2) It is a question how far the streaks of brilliance found

occasionally in mulatto children may interfere with the work of standardization in such a school population. Of course large numbers will overcome this, but it is a distinctly different element from any found in a white school population.

(3) Unquestionably the greatest source of error, and the greatest factor in producing this disparity of results is the *uncertainty of age* of negro children. Both in the public schools and in the reformatory we find negro children absolutely ignorant, not only of date of birth, or time when birthday comes, but also of the numbers of years of age. Constantly such children and their teachers guess at the age. This ignorance and apparent attempt to deceive in regard to age so impressed us at the Juvenile Negro Reformatory that we came to rely almost entirely upon our own judgment of the age of the boy from what we could observe of his size, growth, and physical development, especially of the secondary sex characters. Facts of this sort observed seemed to us very much more reliable than the statements of the boy. No reliable school records were found in either the Reformatory or in the public schools. A special factor making for the unreliability of stated age in the Reformatory consisted in the fact that boys overstated their age with the hope of early dismissal from the institution.

(4) The more tentative character of diagnosis of feeble-mindedness in the negro child as a result of individual mental examination and age rating, which has already been referred to, constitutes another possible source of this disparity stated for the two races.

The Two Per Cent Strata of Intelligence In White Children

Of the 94 white boys given definite mental age rating as a result of individual examination by the Binet-Simon method, 43 were found to be feeble-minded, and 41 of these had made scores by the group tests falling below the lowest two percentile level of public school children. The other two found feeble-minded were between the two and three percentile level of public school children in their group scores.

Of 13 white boys diagnosed as of border-line intelligence, 10 in their group scores fell below the two percentile level of public school children; two others below the three percentile level; and the remaining one below the five percentile level.

The remaining 38 of these 94 white boys who were diagnosed as not feeble-minded are divided as follows in regard to

their percentile ratings. 24 fell below the two percentile level; 9 others below the three percentile; three others below the five percentile level; and two below the ten percentile level. From this it is evident that all the feeble-minded would be caught in the net of the group tests by adhering to the three percentile level. Only six boys who made scores above this level were seen in individual examination. One of these was diagnosed as of border-line intelligence, and the remaining not feeble-minded. It is likewise evident that practically all the feeble-minded would be found by further sorting over by the individual examinational method those boys scoring in the group tests below the two percentile level.

In the two schools for white girls, 25 girls were called for individual examinations and given mental age ratings. 21 of these were below the two percentile level, the remaining 4 being below the five percentile level. None of these girls rated between the two and three percentile level.

15 of these 21 girls below the two percentile level in group tests were found to be feeble-minded; 3 of border-line intelligence; and 3 not feeble-minded. Of the 4 girls rating as of five percentile intelligence, one was found to be of border-line intelligence, and three not feeble-minded, as the result of individual examinations.

In these three groups of white children numbering 382 in all, 119 scored by the A or the X form of group intelligence rating below the limit of the lowest two per cent of 1012 white public school children.

Had the individual examinations been confined to these 119 children, we would have found 56 of the 58 whom we diagnosed as feeble-minded as the result of individual examination, and 13 of the 17 for whom we made provisional diagnosis as being of border-line intelligence.

Using the group tests as the first screening process, and assuming the two percentile limit to have corralled practically all of the feeble-minded, it is seen we would have saved a considerable amount of effort by this use of the group method of intelligence rating. This two per cent would comprise a considerable number of dullards, and some, perhaps, who are not dull, but for one reason or another would not put forth their maximal effort under the conditions of the group test. Many of these, too, are defective characters,—constitutional inferiors, who constitute most serious problems for their caretakers. These need most careful attention from the psychiatrist. Others of this class will be picked from those making

higher scores by their behavior, and by the reported difficulty they cause the management.

Individual mental examination made in this two percentile group, which in these schools constitutes less than 40% of the population, will clear, apparently, about two-fifths of the individuals from suspicion of being feeble-minded. It will pronounce them able to learn and to adapt themselves to their environments so as to be able to manage themselves and their affairs with prudence, and will, therefore, consider them as constituting no danger to themselves, or to others on account of defects of intelligence.

About one half of the two percentile group of these white children prove to be clearly feeble-minded in the sense of this definition just given. The remaining ten per cent of the group we found it necessary to provisionally diagnose of borderline intelligence. We mean by this that our observation has not been sufficient to enable us to pronounce the child so designated as feeble-minded, or as not feeble-minded. We must leave the definite diagnosis to further experience with the child. In many cases we feel that a longer period of experimentation with him is necessary before a definite diagnosis can be given. In fact, with many children of this class the pragmatic test of experience alone will demonstrate whether or not the individual is simply a dullard, defective in some range of imagination, and entirely capable of managing himself and his affairs with prudence, or whether he is so defective through inheritance that he will have to be looked after by the community in order to keep him straight and to save the expense of his crimes and pauperism.

From this point of view the group tests work out very much as would be expected. Army results show about one and one half per cent of the young adult male population to be feeble-minded. It is usually conceded that about one per cent of the total population are feeble-minded in the sense that they have not intelligence enough to manage themselves and their affairs with prudence, and hence need community care to save the expense of their own short-sightedness. If these group tests really measure the intelligence of the children, it should be possible to find all of those who are so defective as to need community care within the group of the lowest two percentile level of public school children. Finer work in the future with group tests may possibly bring us yet nearer to an ultimate sorting out of the feeble-minded by the group method of examination alone.

Without doubt the group intelligence ratings of negro chil-

dren can be made much more serviceable when a means is found of ascertaining definitely the ages of children examined.

SUMMARY

(1) Group Intelligence Rating scores of white boys, white girls, and colored boys, of Alabama Industrial Schools are stratified by the percentile achievements of 1012 white children and 765 colored children of the schools of Jackson, Mississippi.

(2) Only 8 of the 307 white boys are up to the expected school grade. Only 30 others are as little as one year retarded.

(3) The following percentages score below the 50% level of public school children: White boys 86%, white girls 99%, colored boys 84%.

(4) The percentages below the ten percentile level for the three groups in the above order are 64%, 52%, and 32%.

(5) The percentages below the five percentile level for the three groups in the above order are 49%, 44%, and 10%.

(6) The feeble-minded white boys and white girls, and those provisionally diagnosed as of border-line intelligence are practically all found among those scoring below the lowest two per cent of public school children.

(7) This shows the group test to be a labor-saving device in such mental surveys of institutions. Three-fifths of the lowest two percentile scores are feeble-minded or of border-line intelligence in the three schools for white children surveyed.

(8) A larger percentage of negro boys had to be examined individually in order to catch the feeble-minded, and these boys' scores in the group tests range up to the thirty per cent level of intelligence of colored children in the public schools. In fact the twenty per cent level in colored children seems to correspond closely to the two percentile level in white children as a group score limit of feeble-minded intelligence.

(9) Ignorance of actual age in colored children may be responsible for this disparity of results.

(10) Uncertainty of the applicability of Binet-Simon standards to the mind of the negro child may constitute another reason for this disparity in the relations of group and individual intelligence examination results in the two races.

(11) Numbers and percentages diagnosed feeble-minded for the three groups are as follows: White boys 44 (14%), white girls 16 (21%), colored boys 57 (22%).

COMPARISON BETWEEN TWO SCALES FOR THE ESTIMATION OF INTELLIGENCE

By Harry Dexter Kitson, Indiana University

In view of the prevailing interest in methods of estimating abilities, the writer undertook a comparative study of two varieties of scales for the estimation of intelligence.

The first, devised by Pearson, has apparently not been used in America except at The University of Chicago where it is a part of the program of psychological investigation which the writer carried on for a number of years.[1] Some data relative to the use of the scale in this connection have been presented by Ruml.[2]

The second scale was first used[3] in connection with psychological investigations upon the ability of salesmen. The idea formed the basis of the Army Officer's Rating Scale and is now being used as the basis of several other rating scales in process of development.

The Pearson scale (hereafter designated, Scale I) is described at length in *Biometrika*,[4] but lest the files of that journal be inaccessible to some readers a summary will be given here. The scale consists of a number of classes or degrees of intelligence which were defined by Pearson substantially as shown in the form below. On the basis of this classification a number of headmasters and teachers in English schools were requested to rate their pupils. The persons rated included children in grammar school and students in Oxford and Cambridge, totalling about 5,000.

Assuming that the distribution of intelligence follows the curve of probability, Pearson plotted the estimations along this curve; found the numerical value for each class by determining the point on the scale corresponding to the mean value; determined the limits by finding the deviations from the mean

[1] Kitson, H. D. The Scientific Study of the College Student. *Psychol. Rev. Mon. Sup.* 1917, Vol. 23, No. 1.

[2] Ruml, B. The Reliability of Mental Tests in the Division of an Academic Group. *Psychol. Rev. Mon. Sup.* 1917, Vol. 24, No. 4.

[3] "Aids in Selecting Salesmen," published by Bureau of Personnel Research, Carnegie Institute of Technology, November, 1916.

[4] Pearson, K. Relationship of Intelligence to Size and Shape of Head. *Biometrika*, Vol. V., 1906, pp. 105-46.

(measured in terms of the S. D.) required to include in each class the observed frequency.

The S. D. was approximately 100, and so the range of the two classes marked "Fairly Intelligent" and "Slow Intelligent" was divided into 100 parts called *mentaces*. Pearson called average intelligence Zero, and expressed standings above and below by plus and minus scores. The numerical values of the classes were found by determining the number of cases falling between the averages for each two classes (based on the probability integral). These values are shown in Table I. This table differs from the original one of Pearson's, in placing average intelligence at 300 and indicating deviations above and below, by steps; thus eliminating the use of plus and minus signs to indicate deviation above and below the average. It also contains intermediate steps between the mid-points of Pearson's classes allowing finer discrimination for such judges as care to use them.

TABLE I

Numerical Values of Classes and Sub-Classes			
M+471	O+337	Q+220	S+96
M 451	O 322	Q 192	S 62
M—416	O—302	Q—177	
N+391			
N 371	P+282	R+157	
N—353	P 262	R 130	
	P—242	R—116	

At the close of each quarter of the university year instructors of students in the School of Commerce and Administration are given the blank reproduced herewith, with the request that they rate their students. Each student thus receives three estimations in the course of nine months, thirty-six during his entire course. These are added and averaged, thus giving basis for a rank upon the scale of intelligence.

Classification of Intelligence

In order to standardize judgments on mental ability, the following classification of intelligence has been selected. Please note that the scale covers the range of the population at large from the genius to the imbecile.

On the accompanying blanks, please place the letter standing for the class in which you judge the student's mental capacities to fall. What is desired is a judgment of general intelligence and not of classroom performance. If the in-

dividual seems to be in one class with marked leaning toward another class, judge him to be in the more certain class and indicate the direction by + or −.

Class L—*Genius*

Class M—Especially Able: A mind especially bright and quick both in perception and in reasoning; able and accustomed to reason rightly about things on pure self-initiative, even when the data are novel.

Class N—Capable: A mind less likely than M to originate inquiry, but quick in perception and reasoning.

Class O—Fairly Intelligent: A mind ready to grasp and capable of perceiving facts in most fields; capable of good reasoning with moderate effort. This group comprises, say, one-third of the total population.

Class P—Slow Intelligent: A mind slow generally, although possibly more rapid in some fields, but quite sure of knowledge once acquired.

Class Q—Slow: A mind very slow, but not incapable of progress, granted time and considerable effort.

Class R—Slow Dull: A mind capable of perceiving relationship between facts in some few fields with long and continuous effort, but not generally or without external aid.

Class S—Very Dull: A mind capable of holding only the simplest facts, and incapable of reasoning about or grasping the relationship between facts. This group passes into the mentally defective.

Class T—Imbecile

According to the second method (hereafter designated, Scale II) each judge makes his own scale. He chooses from among university students whom he has met, five who possess the degree of intelligence denoted by the five steps on the scale. Then he compares each person to be judged, with each member of the scale and assigns him the appropriate numerical value. As heretofore applied, this method has been used in making estimations concerning complex abilities which are subdivided into simpler qualities: ability as an army officer, for example, into physical qualities, intelligence, leadership, personal qualities, general value to the service. Being concerned in the present study only with intelligence, the writer chose this single trait, modeling directions for rating as *closely as possible according to the directions for rating army* officers. The blank given to instructors is reproduced here, and will make further explanation unnecessary.

RATING SCALE FOR USE IN ESTIMATING INTELLIGENCE

This scale is designed to give basis for the scientific estimation of intelligence, defined as: "the general capacity of an individual consciously to adjust his thinking to new requirements; it is general adaptability to new problems and conditions of life."

Highest 15	Low 6
High 12	Lowest 3
Middle 9	

HOW TO USE THE SCALE

From your acquaintances among university students select five that range from highest to lowest in intelligence.

Place the most intelligent one in class rating 15. Place the least intelligent one in class 3. Select one halfway between these two, of about average intelligence; place him in class 9. Select one who ranks halfway between the middle one and the highest and place him in class 12. Select one who is halfway between the middle and lowest and place him in class 6.

Now take the student whom you are rating and compare him with each of the five in your scale. Give him the number of points following the name of the student he most nearly equals. If he falls between two students in the scale give him a number accordingly (e.g., if between Low and Middle give him 7, 7½ or 8).

POINTS FOR SPECIAL ATTENTION

1. Use as your criterion only General Intelligence, disregarding moral and volitional qualities and scholarship record.
2. In making your scale use students who exhibit clearly the degrees of intelligence assigned to them.
3. The names of the highest and lowest must represent extreme cases.
4. Your scale must be evenly graduated. There should be equal distances between the several steps.

For the purpose of this investigation 100 students were chosen who had been rated the greatest number of times since the adoption of Scale I. Their names were then given to the instructors with the request that they rate the students according to Scale II. It should be stated that the members of the faculty are accustomed to making estimations about students, a systematic plan having been in operation even previous to the adoption of the Pearson Scale; and they responded willingly and whole-heartedly to the request.

Results and Conclusions

TABLE II

	Number of Judgments			Average Scores					Correlations with[5]			
	Min.	Max.	Av.	Min.	Max.	Av.	M. V.	% D.	Grd's	P.E.	Tests	P.E.
Scale I	3	20	10.1	257.0	421.0	331.8	36.0	10	.62	.004	.40	.088
Scale II	1	7	3.0	5.2	13.4	9.9	0.91	10	.66	.012	.32	.094

These scales, so dissimilar to the naked eye, containing different definitions of intelligence, and requiring different techniques in manipulation, nevertheless give results of great similarity, as will be seen from an examination of Table II.

The first thing to strike one's attention is the high average intelligence imputed to this group of university students according to both scales. In terms of Scale I the average for the population-at-large is 300 mentaces; that for these students, 331.8. According to the construction of Scale II the average for *all university students* is 9; that for these students, 9.9. The facts are shown graphically in Graphs I and II, representing the distributions of ratings made by the two scales. The curves both skew toward the upper levels of intelligence. That for scale II is the more nearly normal, probably because the norm taken for the judgments was the university student of medium intelligence, while the norm taken for Scale I was the average intelligence of the population-at-large. The data give good indication that both scales are able to differentiate a university group from a non-selected group.

As to the consistency of the two scales in arranging the members of this selected group with respect to each other, the index of correlation between the two series of ratings is .61, ± .042, certainly significant, though not remarkably high.

As to the respective values of the two scales in measuring intelligence, we have only indirect light. The only access we have to this elusive trait is that gained through its expressions. Two commonly accepted expressions available for this investigation were academic grades and combined scores made in the psychological tests given to each student in his freshman year.[6] Ratings made with Scale I correlate with grades by .62, ± .004; ratings made with Scale II correlate with grades by about the same amount. Ratings made with Scale

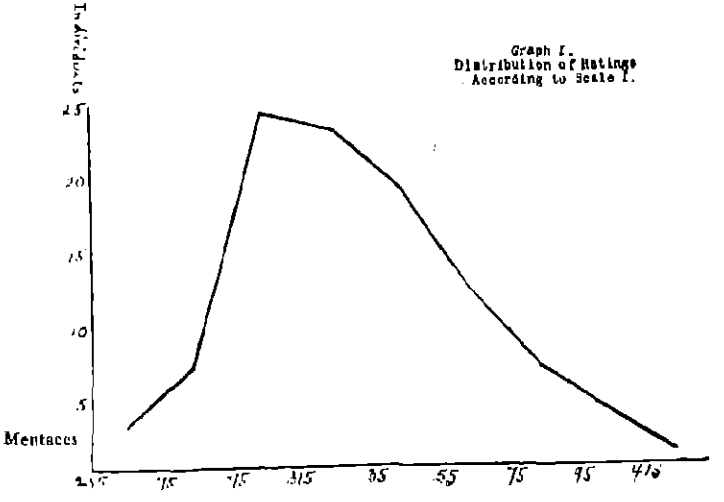
[5] Correlations computed according to formula for adapting the Pearson method to rank-differences.

[6] *loc. cit.*

I correlate with scores made in the psychological tests by .40, ±.088; ratings made with Scale II correlate with test-scores by a closely similar amount. These figures tell us nothing about the *relative* accuracy of the scales in measuring intelligence. They inform us merely that rating by Scale I is about as effective in approximating academic grades as is rating by Scale II; and is about as effective in approximating scores made in the psychological tests as is rating by Scale II.

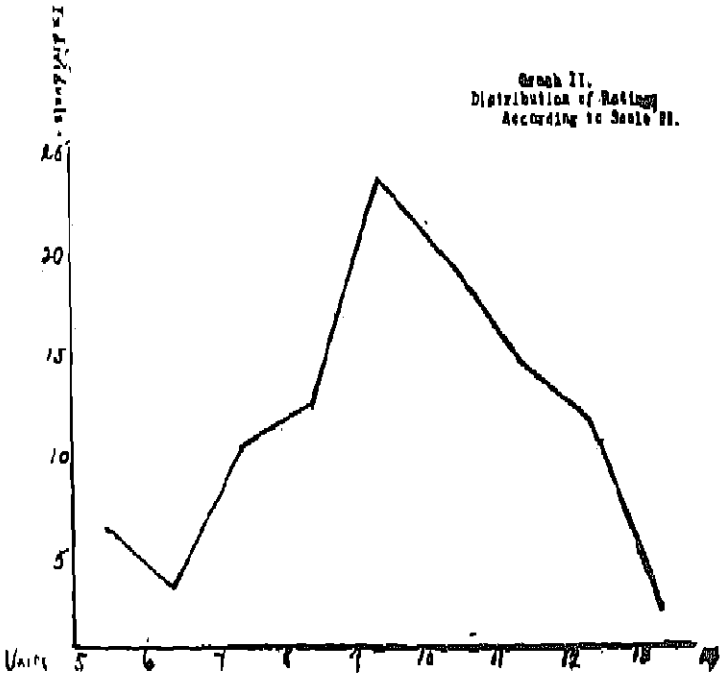
Both series of ratings correlate more closely with academic grades than with psychological test scores. This may argue that in rating, the instructors used the kind of ability leading to scholastic attainment as their chief gauge of intelligence. This conclusion is favored by the fact that when the two series of ratings are combined, the resulting series correlate with grades by .76, ± .028; with tests, by only .39 ± .089.

Both series of ratings agree with the available criteria of intelligence to about the same extent. Does this consistency mean that there is but one thing measured by the two scales even though it is given different definitions and different unit-steps? Or does it mean that the instructors made the same scale out of both? That in spite of their efforts to use a different technique in the two cases, they unwittingly employed the same? The writer is prepared to view the latter as a plausible contingency. It is possible that in making such estimations each judge prefers his own fixed method of



procedure and leans toward it in spite of his efforts to adapt himself to the requirements of one scale or another.

These similarities in the operation of two apparently dissimilar scales suggest that in these days when rating scales are springing up so abundantly, time should be taken for *comparative* investigations. It is quite likely that two methods which at first blush seem widely divergent may turn out upon statistical analysis to be quite similar after all.



THE AVERAGE MENTAL AGE OF ADULTS

By Edgar A. Doll, Fellow in Psychology, Princeton University

Three years ago Professor Lewis M. Terman came to the conclusion that the average mental age of adults is approximately 16 years. This conclusion is still generally accepted and still determines the calculation of intelligence quotients. Terman based his conclusion on the fact that the median mental age of a fair number of supposedly unselected adults proved to be 16 years by the Stanford revision of the Binet-Simon scale. Terman made no attempt logically to explain the fact nor to elaborate the various implications of his conclusion. Arthur S. Otis, however, by logical analysis and by the application of results obtained by Terman and Kuhlman, showed that the curves of distribution of intelligence at different age groups tended to approach each other more and more closely until the curves of distribution for ages above 16 practically coincided with the 16-year curve. Otis, therefore, concluded for logical reasons, making use of the best data available, that 16 years was the upper limit of intelligence development on the average. Strangely enough, these conclusions aroused comparatively little discussion, but were accepted practically at face value. Many psychologists were surprised that the average intelligence of adults proved to be so low, but those who experimented for themselves, came rather generally to the same conclusion as did Terman. A few isolated studies showed that the intelligence of different social classes and of different occupational groups varied rather definitely, but no attempt was made to obtain any large number of really unselected adults, made up of the representative proportions of the different social levels that are found in society at large.

Psychological examining in the Army, however, has indicated that for unselected adults, as judged by draft recruits, a mental age of 16 years is in all probability much too high for the American average adult, at least as he was observed in the Army. The results of the application of group tests at different Army cantonments indicated wide differences in intelligence according to geographical sections of the country. They indicated very important influences of selections, even in groups of over ten thousand. It was also found that occupa-

tional groups were of strikingly different levels of average intelligence. Furthermore, such studies as were made of foreign-born and negroes, indicated very striking differences according to nationality and race.

Perhaps it is too early to come to any general conclusions regarding the significance of the Army data. Perhaps it is impossible for any one man at any one or even several camps to come to a final conclusion, because of the differences just mentioned in the selected nature of supposedly unselected groups. The fallibility of statistics based even on thousands of cases was amply demonstrated in the Army examining. In spite of highly objective methods, which were almost totally independent of any personal equation of examining and scoring, the different camps reported strikingly different results, and even the same camps reported striking variations. At Camp Dix, for example, the average intelligence scores of successive draft quotas became progressively lower during the period of January to August 1918, although apparently there was no good reason why they should do so. For these and for other reasons, it is advisable to wait until the Surgeon-General of the Army or the National Research Council presents the official results of psychological examining in the Army. The present writer, however, has found so much demand for statements of the Army results, and so much need for revision of former concepts, that he feels it is now very desirable to present some sort of statement, even if it be only preliminary, regarding the average intelligence of adults as seen in the Army. In so doing he will present only very general conclusions. In his own work on return to civil life, the writer has found these conclusions of very practical value, and substantially correct. He confesses, for one, that his own previous opinions of the average intelligence of adults were very much distorted by limited experience and by failure adequately to weight the various experimental groups on which his previous conclusions had been based. The adult population of the State of New Jersey, for example, includes such heavy proportions of foreigners and unskilled laborers that the supposedly unselected groups previously observed by the writer now seem to have been very definitely selected in favor of superior intelligence on the basis of social standing.

The Intelligence of Army Recruits

The following table throws some light on the average intelligence of adults in terms of mental age, and compares native whites, foreign-born, and negroes:

MENTAL AGES OF ARMY RECRUITS BY NATIONALITY AND COLOR

Army Letter-Grade Rating	Equivalent Mental Age	Native White %	Foreign born %	Negro %	Army as Whole* %
Illiterate		17	72	62	30
D—	under 9.5	5	55	39	10
D	9.5—10.9	15	29	28	15
C—	11.0—12.9	21	8	16	20
C	13.0—14.9	27	7	10	25
C+	15.0—16.4	19	1	4	17
B	16.5—17.9	9		2	9
A	18.0—19.5	4		1	4
Total Number		9,500	3,700	3,300	1,500,000
Lower quartile (approx.)		11.5	below 9.5	below 9.5	11.0
Median Mental age (approx.)		13.5	9.5	10.0	13.5
Upper quartile (approx.)		15.5	10.5	12.0	15.5

*Observe that the percentages in this column correspond closely to the weighted average percentages obtained by combining the three preceding columns.

The column headed "equivalent mental age," is obtained from the *Basis for the Assignment of Letter Grades,* Appendix A, of the *Examiner's Guide for Psychological Examining in the Army* (Page 91). The column headed "native whites" is based on approximately 9,500 men of the July draft at Camp Dix, New Jersey. The column headed "foreign-born" is based on approximately 3,700 foreign-born men chiefly southern Italians, Russians and Russian Poles from the July draft at Camp Dix, New Jersey. The data of these two columns are embodied in a report from Lieut. Chas. L. Harlan to the Surgeon-General of the Army, entitled "A Study of Foreign-born Men in the July Draft." The column headed "negro" is based upon approximately 3,300 negroes of the July draft at Camp Dix, New Jersey. The data of this column were obtained from a report to the Surgeon-General entitled "A Study of the Intelligence of Negro Recruits," by Lieut. Edgar A. Doll with the assistance of Lieut. Arthur Kornhauser. The column headed "Army as a whole" is obtained from data presented on page 5 of *Army Mental Tests,* Washington, D. C., November, 1918.

The significance of the table is obvious. The median men-

tal age of the total Army proves to be about 13.5 years. This median presumably includes native whites, foreign-born, and negroes. This median is not substantially different from that obtained for native whites at Camp Dix, but as has been previously mentioned, a median based on any one camp, or on less than many thousands of cases, is not entirely dependable. At Camp Taylor, for example, there were very striking differences between the average intelligence of regiments in the same brigade. We may, however, conclude with a fair degree of safety, that the average of totally unselected adults, regardless of nationality and color, is approximately 13 to 14 years.

The other columns of the table indicate striking differences between native whites, foreign-born, and negroes. The average mental age of native whites is 13.5 years, and of foreign-born and negroes is approximately 10 years.

In considering the average mentality of foreign-born it must be clearly understood that there is no implication that these foreign-born adequately represent the average intelligence of their country. On the contrary it is commonly recognized that the type of immigrant received in this country since 1900 is distinctly inferior to the type of immigrant previously received. In general the foreign-born in this study are probably of inferior intelligence in their own countries. These foreign-born do, however, adequately represent the foreigners now in this country between the ages of 20 and 30.

The column of this table headed "Army letter-grade rating," combines both Alpha ratings and Beta ratings, that is, ratings of both literates and illiterates in group examinations. The first horizontal line of the table shows the percentage of illiteracy and indicates the close relationship between per cent of literacy and average intelligence. The foreign-born and negroes were not penalized because of illiteracy, but rather had the advantage. There was a consensus of opinion among Army examiners that illiterate men of inferior intelligence tended to obtain higher ratings in the illiterate group examination (Beta) than they would have obtained in the literate group examination (Alpha) had they been able to read and write. Compared with the native whites, therefore, the foreign-born and negroes have probably higher ratings than they deserved. On the other hand, it is of course conceivable that illiteracy tends to inhibit the full expression of the innate or potential intelligence of a man who has not had the advantage of schooling and an ordinary social environment. In opposition to this, however, is the conviction of most examiners that

the negroes and foreign-born were adequately rated, and if anything, rated higher than their true intelligence warranted.

We come to the general conclusion, then, that the average intelligence of native whites is approximately 13 years, and that the average intelligence of foreign-born and negroes is approximately 10 years. It should be stated that there are striking differences in the mentality of the different foreign-born groups. The column of this table headed "foreign-born," is based chiefly upon southern Italians, Russians, and Russian-Poles, who constitute approximately 2,500 of the total 3,700. It is not feasible at this time to consider the intelligence of the foreign-born in relation to specific nationality. This was an important contribution in the study made by Lieut. Harlan quoted above. Similarly among the negroes there were striking differences between the northern and southern negroes. The southern negroes were decidedly inferior to the northern negroes. The southern negro had an average mental age of about 9 years as compared with the average mental age of about 11 for northern negroes, as indicated in the study of negroes referred to above.

The Age of Average Mental Arrest

The conclusions drawn from the preceding argument indicate that the average intelligence level of native white adults is approximately 13 years. This implies that growth in general intelligence is complete at the life age of 13 years, but this is not a necessary corollary. Such a conclusion cannot be made until successive examinations over a period of years on large numbers of cases, indicate at what age intelligence actually does cease to develop. It is conceivable that while the average intelligence level of adults is 13 years, this development may not have been fully attained until perhaps considerably after the life age of 13 years. There is indeed some reason to believe in a decreasing rate of intelligence growth, so that complete intelligence development may not be attained until an indefinite age.

Some light is thrown on this problem by applying the Army Alpha test to juveniles. The writer has examined 514 children in the Franklin School, Trenton, N. J., with the Alpha group test. The median Alpha scores at each age are as follows:

Median Age	9.5	10.5	11.5	12.5	13.5	14.5	15.5
Median Score	26	37	61	81	85	80	88
No. Cases	29	38	63	96	139	75	22

These figures indicate that the median intelligence of children above 12 years does not significantly exceed the median of 12-year-old children. There is a slight superiority of 13 over 12, but after this there is a decrease. These children were examined in the eight grammar grades of a typical average public school. An analysis of the age and grade distribution indicates that the children are typically average school children. The children in the 15-year-old group might, however, be considered as slightly below average since they are slightly over age for the eighth grade.

A similar study has been made of 480 delinquent boys at the State Home for Boys at Jamesburg, N. J. The median scores at each age are as follows:

Median Age	9.5	10.5	11.5	12.5	13.5	14.5	15.5	16.5
Median Score	6	8	15	20	31	34	30	29
No. Cases	15	33	51	82	91	102	73	28

It will be observed that at each age these Alpha scores are markedly below the scores of the public school children of the same age. In fact, the average inferiority amounts to about 3 years. The point at issue, however, is that here again average development is complete at about 13 years, since the median scores of ages above 13 do not materially exceed those at 13, and in fact at some ages are below. This substantiates the result found in the public school and indicates that on the average, mental development is complete at about 13 years.

It is not possible to cite any satisfactory statistics for any long period of years on the successive mental ages of normal children of adolescent age. The writer is, however, conducting a study of approximately 250 feeble-minded children who have been re-tested over a period of at least four years each. It is too early to state final results, but as far as the study has progressed it is evident that the feeble-minded do not develop in intelligence after the life age of about 13 years. Contrary to previous beliefs, the age of final arrest is not a function of mental retardation, but is a fairly definite life age. Some feeble-minded of each grade continue to develop in general intelligence, even though only slightly, up to about 13 years of age.

From all of this evidence it appears that not only is 13 years the average intelligence age of adults, but the implication is very strong that the full level of intelligence has been reached at the life age of approximately 13 years. This is not to be understood, however, as meaning that the entire mental growth is attained at that age. Maturity and experience must

certainly continue to effect changes in *mentality* after the age of 13 years, but the intelligence *level*, on the average, remains fixed. This does not mean that no one develops in intelligence after 13 years of age, or that no one has a mental age above 13 years. On the contrary, 50 per cent of these groups have mental ages above 13 years. It is not possible to say at this time which of the 50 per cent continue to develop, nor to what mental age they develop, nor at what life age limits mental growth ceases. The only permissible conclusion from these data is that *on the average,* or for 50 per cent of presumably unselected cases, intelligence growth is practically complete at 13 years.

The writer wishes to emphasize that it would be fatal to both theory and practice to establish anything more than some new hypotheses on the data of this paper. Fifty per cent of these subjects do seem not to have developed intellectually after 13 years, but we cannot say much about the other 50 per cent. Repeated individual examinations on a large scale are now our most pressing need.

Intelligence and Feeble-mindedness

There has been a tendency among clinical psychologists to lay increasing emphasis upon intelligence as a diagnostic symptom in determining feeble-mindedness. This emphasis is entirely proper but has been subject to a certain logical error. It was found by examining mental defectives in institutions that practically no feeble-minded "tested" higher than 13 years in the Binet-Simon scale. It was then concluded both by implication and by actual statement, that all who tested *under* 13 years were feeble-minded. The upper limit of feeble-mindedness was tacitly assumed as the lower limit of normality. Obviously this is the logical error of the converse. If all A is B, the only necessary consequence is, that *some* B is A. For example, if all negroes have mental ages below 15 years it does not follow that all persons with mental ages under 15 years are negroes.

From practical experience, some clinicians have developed the concept of borderlinity. They recognize that there is a certain mental age zone within which the intelligence level alone breaks down as a diagnostic instrument. Within the limits of this zone some persons may be feeble-minded, some doubtful and some normal. In this connection the important contribution of the Army examining lies in the demonstration that persons with mental ages as low as even 7 or 8 years may

be socially normal, and that others with mental ages as high as 13 or 14 years may be socially feeble-minded.

The fundamental conception of feeble-mindedness, of course, is one of social competence. A feeble-minded person is not a person of a particular mental age, but is one who by reason of inferior intelligence is incapable of supporting himself independently of supervision in society, after he has reached the age of social responsibility or physical maturity. The mental age limits which were developed by clinicians in the Army on the basis of practical experience to designate this social responsibility are somewhat as follows:

Diagnosis	Native White	Foreign-born	Colored
	Mental age	Mental age	Mental age
Invariably feebleminded	under 8	under 7	under 7
Usually feeble-minded	8.0— 8.9	7.0— 7.9	7.0— 7.9
Usually inferior normal	9.0—10.9	8.0— 9.9	8.0— 9.9
Usually average normal	11.0—13.9	10.0—11.9	10.0—11.9
Usually superior normal	above 14.0	above 12.0	above 12.0

In this table "invariably feeble-minded" means that persons of the intelligence level specified invariably proved to be feeble-minded when account was also taken of their social success, educability and general mental responsibility. "Usually feeble-minded," means that those persons within the mental age limits specified, usually, or in about 8 cases out of 10, proved to be feeble-minded when other factors are considered. In this group those who were not feeble-minded were ordinarily subnormal or borderline cases. "Usually inferior normal," means that in about 8 cases out of 10 the individual within the mental age limits specified proved to be normal on the basis of social responsibility, literacy, educability and general mental qualities. Those not found to be normal were ordinarily high-grade feeble-minded or borderline cases.

This grouping is not based on statistical or experimental study, but is an empirical classification developed from extensive experience. In the Army it was found impossible to use intelligence alone with any degree of certainty as a diagnostic method above intelligence level of about 8 years. Mental age was an invaluable diagnostic *aid*, however, and was perhaps the most important consideration next to a man's social history. In the Army it was found necessary to weigh social success very heavily in determining normality or feeble-mindedness, and in the mental examinations it was found important to take serious account of the quality of intelligence as well as the

level. Temperamental characteristics were also very important; prudence, forethought, judgment, reason and insight were opposed to mere verbal brightness.

It will be noted that this classification uses lower mental age standards for foreign-born and negroes than for native whites. This implies that the mental ages of foreign-born, negroes, and native whites are of somewhat unequal value in determining social fitness. The mental ages of the foreign-born and negroes seemed to be accurately measured, but due to some unknown psychological quality, the social value of mental age for foreign-born and negroes was somewhat greater than for the same age for native whites.

This empirical conclusion suggests an important problem for experimental psychology. If the prognostic value of mental ages is not constant for the different races, psychology should certainly aim to discover the reason. It seemed desirable in the Army, and still seems desirable, for practical social purposes, to judge individuals according to standards for their race and social class, but this leads to some confusion in the diagnoses obtained. We have come so much to depend upon mental age as an expression of social fitness, that it will be somewhat difficult to employ multiple standards for different social groups.

Average Intelligence and the I. Q.

It is now the practice to use 16 years as the basis of calculation for subjects whose life age is higher than 16. But if the implications here presented should be substantiated by further investigation, then it will become necessary to compute I. Q's. on a 13-year final age. This will make necessary a revision of previous work in which I. Q's. have been stated independently of age. Until the question of the average age of final mental development is settled, it should be urged upon all advocates of the I. Q. always to accompany I. Q's. with the actual ages of the individuals tested. This is important anyway, since I. Q's. are not directly comparable independent of life ages any more than diagnoses by race and class are comparable independently of the race and class standard. An I. Q. of 50 where mental age is 8, cannot be directly compared with an I. Q. of 50 where the mental age is 4.

It will be noted from the intelligence classification given above, that the I. Q. limits of feeble-mindedness are different from those advocated by Terman and his followers. According to this table, the "invariably feeble-minded" are found to have I. Q's. under 50 as opposed to the I. Q. of 70 advocated

by Terman, and the "usually feeble-minded" are found to have I. Qs. between 50 and 55, instead of between 70 and 75, as advocated by Terman. But, if the I. Q's. of the future come to be calculated on the basis of 13.5 years as the average adult intelligence, instead of 16 as now computed, then I. Q's. under 60 will "invariably" denote feeble-mindedness, and I. Q's. under 64 to 70 will "usually" denote feeble-mindedness. Thus we might retain the present I. Q. diagnostic limits advocated by Terman, but they would have to be calculated on a different basis.

Summary

1. The average mental age of adults is found to be approximately 13 years instead of 16 as heretofore believed.

2. The growth of general intelligence is found to be practically complete on the average by 13 years of age and is not on the average thereafter exceeded. This applies strictly to the level of intelligence or degree of brightness as opposed to intelligence plus maturity, experience, and acquisitions.

3. These conclusions are induced by four separate investigations on totally different types of subjects and from different points of view, namely:

(a). The application of Alpha and Beta group intelligence tests to about 1,500,000 soldiers and recruits, where the average mental age is found to be about 13 years.

(b). The application of Alpha group intelligence tests to about 500 typical public school children, where the median scores of ages above 13 do not exceed the median for 13 years.

(c). The application of Alpha group intelligence tests to about 500 juvenile delinquent boys, who are found to be of inferior intelligence but whose level of intelligence is fully attained by 13 years, on the average.

(d). The repeated application of Binet intelligence tests to about 250 feeble-minded persons of wide ranges of life age and mental age, who individually may develop in intelligence up to a life age limit of 13 years and not thereafter.

4. The conclusions are induced not merely by the empirical data of each separate investigation, but particularly by the consistency of agreement obtained from such strikingly different methods of approach.

5. Striking differences are observed between the average levels of general intelligence of native whites, foreign-born, and negroes. Striking differences also are found for different social classes of native whites, different nationalities of foreign-born, and different geographical groups of negroes.

6. A logical fallacy is exposed in the application of mental age limits to the diagnosis of normality and feeble-mindedness. The conception of borderlinity in relation to mental age is much extended.

7. The calculation of intelligence quotients is found to be seriously disturbed by the uncertainty of the exact life age at which intelligence development is to be considered complete on the average.

8. The standardization of mental age scores for ages above 13 years by any method other than the percentile scores is not attempted, but is found to be impeded by the difficulty of obtaining fully representative unselected groups of subjects outside the grammar schools.

Addendum

It is now possible to strengthen the above argument still further. Nearly 550 delinquent youths and young men between the ages of 16 and 29 inclusive, inmates of the New Jersey Reformatory, have recently been examined by Mr. J. W. Ellis, Assistant Psychologist to the N. J. Dept. of Institutions and Agencies, by means of the Army group test Alpha. The median Alpha scores at each age are as follows:

Median age	16	17	18	19	20	21	22	23	24	25	26	27	28	29	Total
Median score	59	62	56	58	57	43	28	69	38	39	60	98	22	38	56
No. cases	68	115	109	74	47	40	17	15	8	11	11	6	8	5	541

It is obvious from this table that there is no consistent increase in score with increase of age. During the range of years where there is a sufficiently large number of cases for the median to be dependable (16 to 21) there is no significant variation in score except for an unaccountable drop at year 21. The variations beyond year 21 may be accounted for on the basis of selection and the small number of cases. The fact that there is no consistent trend in these variations is, however, more significant than the variations themselves. The median score for the group represents a median mental age of approximately 14 years. The group is a selected group on the basis of delinquency yet the average intelligence score is nearly the same (slightly higher) than the average score of army recruits. Here again, however, we are not so much interested in the absolute score as the fact that there is no steady increase in score within increase in age. The intelligence levels above 16 are not superior to that of 16. It is probable, in the light of the public school and the Jamesburg Home

median scores that the 16-year median itself is perhaps not greater than it would have been had we had age groups extending down to 13 or 14 years, but this is conjectural.

This reformatory group supplies a fifth group and a new type of subject. The group covers the age-period between the public school group and the Army adult group. It fills the gap and gives a continous range of ages from 6 to 31 years. (The scores for years 6, 7 and 8 years have been omitted in this paper.) The continuity is, of course, broken by the variations in types of subjects, but an adequate basis of preliminary argument is afforded in the actual results. The new data confirm the preceding argument that there is no increase in average general intelligence scores after a chronological age of presumably 13 or 14 years. Again we must insist that the scores of individuals range far above the 13-year level, with the presumption, however, that such scores represent superior deviations of the 13-year group and not age-increases as such. These superior scores also demonstrate that the 13-year level is not caused by limitations in the scale of tests itself, since individuals do score far beyond the median, thus proving the practical possibility of obtaining high scores if the capacity is present in the individual.

A PSYCHOLOGICAL INVESTIGATION OF THE LIKELIHOOD OF CONFUSION BETWEEN THE WORDS "COCA-COLA" AND "CHERO-COLA"[1]

By RICHARD H. PAYNTER, JR.

This is a report of a series of psychological experiments conducted to investigate the desceptive similarity or likelihood of confusion in visual appearance and in sound between the word trade-marks "Coca-Cola" and "Chero-Cola." The report also compares the likelihood of confusion between these words with that found to exist between word trade-marks that have been held in judicial proceedings to conflict and with others that have been held not to conflict. Two distinct methods were employed in the investigation: the recognition or identification method, and the method of relative position. Four separate experiments were conducted, three with recognition and one with relative position. The recognition experiments were on confusion in visual appearance, and the relative position experiment was on confusion in sound.

A. VISUAL RECOGNITIVE CONFUSION

1. *General Procedure and Technique.*—The visual recognition or identification method was used because it involves the same mental process as that employed in daily life by the ordinary purchaser in buying a bottle of "Coca-Cola." In ordinary transactions of business where the article is sold across the counter, the purchaser examines the article to see if it is what he wants, or whether it is the same as that which

[1]This study is a practical application of simple, straightforward psychological methods to an actual trade-mark case. The study was used in a case by The Coca-Cola Company in an opposition to the word "Chero-Cola."

A more extensive investigation will appear in a forthcoming number of the Archives of Psychology, entitled "A Psychological Study of Trade-mark Infringement."

The writer expresses his deep gratitude to Professor Edward K. Strong, Jr., of the Carnegie Institute for Technology, for the kind advice and help he has given in devising the experiments.

he has known or bought before. In deciding this matter, the purchaser is guided by the recognition or identification of the article as a whole, by certain features of it, or by a combination of both these methods. It often happens, however, that a prospective purchaser has heard or seen only the word trade-mark, but not the article itself nor its container. In such cases he has to rely on his memory of just the word trade-mark in identifying the article. Likewise, the observer or individual in the experiment has to rely on his memory of the word trade-mark in order to tell whether or not what is before him now is the original trade-mark that he has just seen. The task assigned to him in the experiment was simply to indicate whether the words before him were just shown to him.

The individual who submitted himself to do the recognition experiment was shown in the presentation 20 slips of paper on each of which was typewritten a word trade-mark, or a word trade-mark and the name of its commodity. These slips were presented one at a time at the uniform rate of one a second, by the writer who conducted all the experiments. Fifteen seconds after the exposure of the last slip in the presentation, the test set of slips was given to the observer. The test consisted of 40 slips, and from it the observer was asked to pick out those that he recognized as having just been seen and those that had not been seen. There were two variations of the formation of the test, according to the problem of the experiment. When the problem was to determine the confusion between the words "Coca-Cola" and "Chero-Cola," "Coca-Cola" appeared in the presentation, whereas the imitation "Chero-Cola" appeared in the test together with 19 trade-marks which were duplicates of 19 shown in the presentation, and 20 trade-marks which were new, i. e., they had not been shown in the presentation. When the problem was to compare the confusion of the imitation "Chero-Cola" with that of 9 other litigated imitations, "Coca-Cola" with their 9 originals appeared in the presentation, whereas "Chero-Cola" appeared in the test together with these 9 imitations, 10 duplicates of the remaining 10 trade-marks shown in the presentation, and 20 new marks. In all these experiments "Coca-Cola" was shown in the presentation, but in the test where it did not appear the observer had to decide whether he had seen "Chero-Cola" which was not shown in the presentation. If he decided that he had seen the imitation he was confusing it with the imitation; if he decided that he had not

seen the imitation he was not confusing it with the original.[2] Where the comparison was made with other litigated imitations the name of the article or commodity appeared with each trade-mark. There are two experiments where no comparison was made, one showing the trade-marks without, and the other with, the names of their articles. It should be stated here, however, that although the words in the recognition experiments entered the mind of the observer through vision, not only did similarity in visual appearance, but also in sound, linguistic formation, and meaning or significance contribute to confusion.

The following directions were given to each observer before he did the experiment:

"You are going to be shown one at a time a number of ordinary word trade-marks like 'Uneeda,' 'Garford,' and 'Celluloid.'" (In the experiments where the trade-marks were used with the names of the articles, the observer was told of this and the examples were given as "Uneeda Biscuit," "Garford Automobile," etc.) "You are to read all the words on each slip of paper. Read them naturally as though you were reading an advertisement in a magazine or in a street car. Fifteen seconds after you have been shown the last mark, you will be given a second set (the test) and asked to pick out those marks you have just seen in the presentation and those which you have not seen. You will be further asked to sort the marks into six piles, according to the degree of your confidence or certainty of your recognition of your marks. There are three degrees of certainty for the marks that are recognized as seen, and three similar degrees for those that are recognized as not seen. The three degrees are 'absolutely certain,' 'reasonably certain,' and a 'faint idea.'" (After the observer has indicated his recognitions of all the marks in the test, he was asked how familiar he had been with the word "Coca-Cola"). "Would you say you had been 'perfectly familiar,' 'moderately familiar,' 'just familiar,' or

[2] These two statements are not strictly accurate as they do not take into consideration the possibility of the presence of a small per cent. of observers who would not have been able to recognize "Coca-Cola," and in whom the perception of "Chero-Cola" would not recall "Coca-Cola." These observers would be guessing in recognizing whether or not they had seen "Chero-Cola," and would guess correctly in about 50 or 60 per cent. of the cases, thereby influencing equally the correct and incorrect recognitions of "Chero-Cola." As the 10 original trade-marks are approximately equally well recognized or familiar, accounting for this factor would introduce no significant change in the data or conclusions.

'unfamiliar' with the word 'Coca-Cola' before you did this experiment?"

Thus, three important facts were obtained from the experiment on each observer. The first fact was, whether or not he confused "Chero-Cola" with "Coca-Cola." The second fact was, how confident he was that he was correct, or whether he was in doubt about his recognition. And the third fact was, how familiar he had been in daily life with the word "Coca-Cola." The question of visual confusion will be determined then by the data from these three sources of information.

Each observer did only one experiment; and he did not do the same experiment more than once. The observer was not informed of the purpose of the experiment, nor did he know beforehand that an original trade-mark would be replaced by an imitation in the test. Furthermore, nothing was done to cause the observer to have any suspicion that a substitution had been made. The task set the observer in the experiment was not arduous, nor unpleasant. The experiment was done with the observer's own free will. No prize or reward was offered for the highest individual record. After each observer did the experiment the marks in both the presentation and test were thoroughly shuffled. As their positions in both series were thus determined by chance, no trade-mark was given undue prominence by its position.

Ninety ordinary individuals were employed as observers in the recognition experiments. They represented more than 60 different professions and occupations. Their ages were generally between 20 and 45. These individuals represent a random picking from ordinary individuals that resided in New York City, in the month of July, 1915. No effort was made to obtain exceptionally bright nor exceptionally dull individuals.

All the trade-marks or trade-names studied in the experiments are used or have been used in commerce. When the names of the articles are used in the experiments in connection with the trade-marks, they are the names of the same articles or commodities on which the trade-marks have actually been used. The names of the goods represent many classes of goods of various descriptive properties. In the presentation no trade-mark nor name of a commodity was the same as any other. The names of the commodities of the original trade-marks and their respective imitations were identical. The 20 new trade-marks and the names of their commodities were all different from the others, either in the test or in the presentation.

The trade-marks and names of the commodities were all typewritten in the same kind of type. The experiments did not reproduce the style, coloring, or size of the trade marks or names of the commodities as used on the goods or in advertisements. The color of the ink used was black; and great care was taken in typewriting the words to keep the blackness of the ink the same for all the letters. All the words were typewritten on the same kind of white paper. Each slip of paper was two and three-quarters by four and one quarter inches in size. When the name of the commodity was used it appeared on the next line directly beneath the trade-mark. All the words appeared in the second horizontal quarter from the top of the slip. The four following reproductions are samples of some of the slips, which are identical with those actually used, with the exception of being reduced one-fourth.

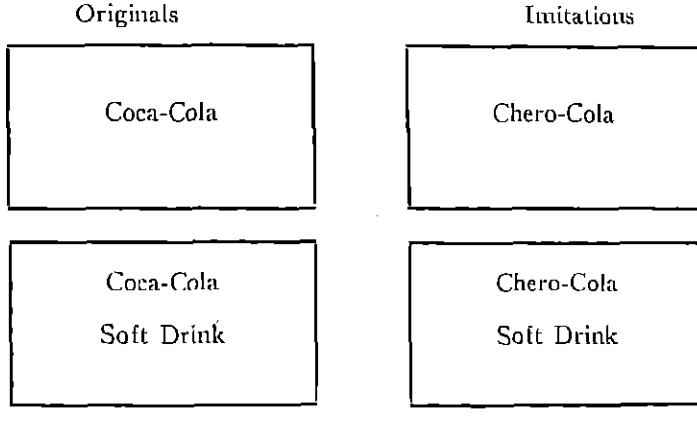


FIG. 1

2. *Results.* a. Experiment in which Appeared One Imitation, the Names of the Articles not being Shown with the Trade-Marks.

The word "Coca-Cola" and 19 other trade-marks appeared in the presentation series; the test series included the word "Chero-Cola," 19 duplicate trade-marks and 20 new trade-marks.

TABLE I

THE NUMBER AND PER CENT. OF 40 OBSERVERS THAT CONFUSED THE WORD "CHERO-COLA" WITH "COCA-COLA"

Number of Observers	Trade-Marks		Confused	
	Original	Imitation	Number	Per Cent
40	Coca-Cola	Chero-Cola	11	28

According to Table I, approximately one out of every three observers or 28 per cent. actually takes the word "Chero-Cola" to be "Coca-Cola." Of the 11 deceived by "Chero-Cola" 9 were "absolutely certain," 1 "reasonably certain," and 1 had a "faint idea" that they had seen it. Thus most of the observers mistaken were quite confident that they were correct. Moreover, 9 of those confused by "Chero-Cola" were "perfectly familiar," and 2 "moderately familiar" with the original trade-mark "Coca-Cola." It is seen from this that observers even very familiar with the word "Coca-Cola" were deceived by the imitation "Chero-Cola."

Throughout this investigation the figure representing the per cent. of individuals in a group confused by an imitation is not absolute. That is, it does not mean that just this per cent. of observers under the multitudinous varying conditions of everyday life will be confused. The percentage of confusion, is, however, a scientific measure depending on the observer's memory of the original mark and its similarity to the imitative.

b. Experiment in which Appeared One Imitation, the Names of the Articles being Shown with the Trade-Marks.

The problem of this experiment is to determine the likelihood of visual confusion between the phrases "Coca-Cola Soft Drink" and "Chero-Cola Soft Drink." There is no other difference in experimental technique between this experiment and the previous one, except that here each mark is used in connection with the name of its commodity. Table II. gives the number and per cent. of 25 observers that confused the words "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink."

TABLE II

THE NUMBER AND PER CENT. OF 25 OBSERVERS THAT CONFUSED THE WORD "CHERO-COLA" WITH "COCA-COLA," WHEN APPLIED TO THE WORDS "SOFT DRINK"

Number of Observers	Trade-Mark Original	Trade-Mark Imitation	Name of Article	Confused Number	Confused Per Cent
25	Coca-Cola	Chero-Cola	Soft Drink	17	68

Seventeen or 68 per cent. of the group confused "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink." When used in connection with the words "Soft Drink" 2 out of every 3 observers take "Chero-Cola" for "Coca-Cola." Of the 17 deceived by "Chero-Cola" 12 were "absolutely certain," 2 "reasonably certain," and 3 had a "faint idea" that they had seen it. Moreover, 14 of those confused were "perfectly familiar," 2 "moderately familiar," and 1 "just familiar" with

the original mark "Coca-Cola." As in Table I. the observers in Table II. were nearly all quite confident of their false recognitions of the imitation, and also very familiar with the "Coca-Cola."

The figures in Table II. show about two and one-half times as much confusion as do those in Table I. It is the addition of the name of the commodity to both the original and imitative marks that furnishes an explanation for this increase. As the observers in the present experiment had to read in the presentation the words "Coca-Cola Soft Drink" in the same time that those in the previous experiment had to read just the word "Coca-Cola," the word "Coca-Cola" in the former not being so well perceived, was not so firmly established as in the minds of those of the latter group. Having to contend with a weaker mental impression of "Coca-Cola" confusion with "Chero-Cola" was more likely. Furthermore, as the two perceptions "Coca-Cola Soft Drink" and "Chero-Cola Soft Drink" are relatively more similar than the two perceptions "Coca-Cola" and "Chero-Cola," confusion between the former is hence greater.

c. Experiment in which Appeared 10 Imitations, the Names of the Articles being shown with the Trade-Marks.

The problem in the present experiment is to compare the likelihood of visual confusion between the words "Coca-Cola Soft Drink" and "Chero-Cola Soft Drink," with the likelihood of visual confusion found to exist between other pairs of similar word trade-marks that had been the subject of actual adjudication of infringement or non-infringement. Nine legal decisions [3] on the similarity of word trade-marks were selected

[3] Holeproof, Knotair, hosiery; Holeproof Hosiery Co. v. Wallach Bros., 190 Fed. Rep. 606 (U. S. Cir. Ct.).

Grape-Nuts, Grain-Hearts, cereal food; Postum Cereal Food Co., Ltd., v. American Health Food Co., 119 Fed. Rep. 848 (U. S. Cir. Ct. of App.).

Cascarets, Castorets, remedy; Sterling Remedy Co. v. Spermine Medical Co., 112 Fed. Rep. 1000 (U. S. Cir. Ct. of App.).

Gold Dust, Gold Drop, washing powder; N. K. Fairbank Co. v. Luckel, King & Cake Soap Co., 102 Fed. Rep. 327 (U. S. Cir. Ct. of App.).

Listerine, Listogen, antiseptic; Lambert Pharmacal Co. v. Bolton Chemical Corporation, 5 Trade-Mark Rep. 38 (U. S. Dist. Ct.).

Mother's, Grand-Ma's, waxing pad; E. A. Bromund Co. v. Columbia Wax Products Co., 200 O. G. 1115 (Ct. of App., D. C.).

Sozodont, Kalodont, tooth paste; K. K. Landespriv et al. v. Hall & Ruckel, 1911 C. D. 329, 165 O. G. 732 (Ct. of App., D. C.).

Green River, Green Ribbon, whisky; Lang v. Green River Distilling Co., 1909 C. D. 476, 148 O. G. 280, 33 App. D. C. 506 (Ct. of App. D. C.).

Club, Chancelor Club, cocktails; In re S. C. Herbst Importing Co., 1908 C. D. 383, 134 O. G. 1565, 30 App. D. C. 297 (Ct. of App. D. C.).

for comparison. Five were adjudications of infringement, in which the use of the imitative marks was enjoined, and four were adjudications of non-infringements in which injunctions against the use of the alleged illegal imitations were refused. In selecting these decisions two principles were followed; first, that the original mark in the decision should be widely known; and second, that no other circumstance or reason than that of similarity between the word trade-marks should have determined the point of infringement or non-infringement in the decision.

The first principle was set up because the word "Coca-Cola" is widely known, and is pretty familiar to a great many individuals in New York City, where all the experiments were conducted. And furthermore, because an imitation would probably cause less confusion in the case of a more familiar original mark than in the case of a less familiar original, a comparison of the confusion caused by their respective imitations would unfairly show relatively greater confusion for the imitation of the less familiar original. Thus, the confusion brought about by the imitation of "Coca-Cola" will be properly compared with the confusion brought about by the infringing and non-infringing imitations, if their originals are approximately as well-known as "Coca-Cola." It should be stated here, however, that not all the decisions studied have originals as well known as "Coca-Cola." These marks will receive separate treatment.

The second principle was set up for two reasons. The first was that, the experiment measures visual confusion of just word trade-marks. The second was that, if various factors such as the question of the validity of the trade-marks, unclean hands, similarity of the type, color, or other features of the label or package entered in the decisions and operated to influence the point of infringement or non-infringement, we could not properly compare the experimental results of these complicated decisions with each other, nor with the simple decisions of confusion of just the word trade-marks, nor with "Chero-Cola." It is obvious that decisions determined by confusion of word trade-marks plus certain other reasons are not in the same legal nor psychological categories as those determined simply by the confusion of just the word trade-marks. The decisions in the former categories could not be properly given as authorities governing those in the latter, or vice versa.

Notwithstanding the fact that some selection has been made in obtaining these decisions, no selection was made to obtain

decisions whose trade-marks seemed to have only little likelihood of confusion or less than that of the imitation of "Coca-Cola." That is, no judgment was exercised to set forth as very great the confusion caused by "Chero-Cola" by contrast with imitations that might be expected to show only a slight degree of confusion. In this respect the decisions represent a chance or random sampling of infringements and non-infringements, affording thereby a fair comparison of "Chero-Cola" with infringements and non-infringements of well-known original trade-marks in general. In this respect the experimental method gives a scientific comparison, not one biased in advance to argue a decision of infringement against the word "Chero-Cola."

This experiment is the same as the previous one, except for the presence of 10 imitations in the test series instead of one. The words "Chero-Cola Soft Drink" were in the test series together with 9 other imitations, 10 duplicates, and 20 new trade-marks.

TABLE III

THE NUMBER AND PER CENT OF OBSERVERS CONFUSED BY THE WORD "CHERO-COLA," EACH OF THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS, WHEN APPLIED TO THE NAMES OF THEIR ARTICLES

Decision	Trade-Marks Original	Trade-Marks Imitative	Name of Article	Confused No.	Confused P.C.
N	Sozodont	Kalodont	Tooth Paste	24	48
I	Green River	Green Ribbon	Whisky	23	46
N	Mother's	Grand-Ma's	Waxing Pad	20	40
	Coca-Cola	*Chero-Cola*	*Soft Drink*	*10*	*40**
I	Club	Chancellor Club	Cocktails	17	34
I	Listerine	Listogen	Antiseptic	16	32
I	Cascarets	Castorets	Remedy	14	28
I	Gold Dust	Gold Drop	Washing Powder	13	26
N	Holeproof	Knotair	Hosiery	12	24
N	Grape-Nuts	Grain-Hearts	Cereal Food	9	18

*This percentage is reckoned on the basis of 25 observers in the group; all the other percentages are reckoned on the basis of 50 in the group.

Table III. gives the number and per cent. of observers confused by each of the imitations. The first column in the table shows whether the imitation in the decision was held to be an infringement (I) or a non-infringement (N). The second column gives the original trade-mark in the decision; the third the imitative trade-mark; and the fourth the name of the commodity to which both the original and imitative trade-

marks were applied. The fifth column gives the number of observers confused; the sixth and the last per cent. confused. The marks are arranged in an order of confusion from most to least.

"Sozodont—Kalodont," a non-infringement case, is the most confusing pair, 24 observers or 48 per cent. of the entire group were confused. "Green River—Green Ribbon," an infringement, stands second highest with 46 per cent. confused. "Coca-Cola—Chero-Cola" ties with a non-infringement, "Mother's—Grand-Ma's," for the third most confusing position with 40 per cent. confused. The differences between these four pairs of marks are not large. Of the 10 observers confused by "Chero-Cola" 9 were "absolutely certain," and 1 had a "faint idea" that they had seen it. Nine of those confused by "Chero-Cola" were "perfectly familiar," and 1 "moderately familiar" with the word "Coca-Cola." As in the two previous experiments most of the observers confused were very familiar with the word "Coca-Cola," and they were scarcely even doubtful in their errors. Two imitations are more confusing than "Chero-Cola," one of these is a non-infringing imitation and the other an infringing. Four infringements and two non-infringements have lower scores than "Coca-Cola—Chero-Cola." It is 6 per cent. lower than the most confusing infringement "Green River—Green Ribbon;" and 14 per cent. higher than the least confusing infringement "Gold Dust—Gold Drop." It is 8 per cent. lower than the most confusing non-infringement "Sozodont—Kalodont;" and 22 per cent. higher than the lowest non-infringement "Grape-Nuts—Grain-Hearts."

The original marks in the table that are about as well-known in New York City as "Coco-Cola" are perhaps "Gold Dust," "Grape-Nuts," and "Cascarets." The remaining original marks, with the exception of "Sozodont," contain ordinary words which are met with frequently enough in daily life to be perhaps as familiar as the three above marks. It seems that the combination of the two ordinary words in "Holeproof" puts it also under the latter class of marks. Let us now compare the scores of the imitations of these two classes with the score of "Chero-Cola." It is more confusing than the imitation of any of the most widely known original marks. Two of these imitations are infringements and one a non-infringement. In order of amount of confusion these marks are:

Coca-Cola	*Chero-Cola*	*40*	per cent.
Cascarets	Castorets	28	" "
Gold Dust	Gold Drop	26	" "
Grape-Nuts	Grain-Hearts	18	" "

" Coca-Cola—Chero-Cola " has confused 12 per cent. more observers than " Cascarets—Castorets; " 14 per cent. more than " Gold Dust—Gold Drop; " and 22 per cent. more than " Grape-Nuts—Grain Hearts." Of the less well-known original marks one non-infringement " Sozodont—Kalodont," and one infringement " Green River—Green Ribbon " are more confusing. " Mother's—Grand-Ma's," a non-infringement ties with " Coca-Cola—Chero-Cola," and all the rest are less confusing than it.

It will be observed that in Table III " Chero-Cola " shows less confusion than in Table II. In both experiments the marks were used in connection with the name of the commodity. In Table III the per cent. confused is 40, and in Table II, 68 per cent., the difference being 28 per cent. The reason for the decrease in Table III was that " Chero-Cola " appeared in the test together with 9 other imitations. The fact that the observer was usually confronted with several imitations before the imitation of " Coco-Cola " appeared, was in itself a kind of fore-knowledge and warning to beware of imitations. In the experiment of Table II. the presence of one imitation among 39 different trade-marks in the test, 19 being duplicates of what they had seen and 20 new, would find the observer not expecting to be fooled by an imitation. In the experiment of Table III the presence of 10 imitations in the test together with 10 duplicate and 20 new marks would, on the contrary, act as a warning to be more cautious, wary, and careful in recognizing all marks. The natural result followed this condition; that is, there was a cutting down of the confusion caused by " Chero-Cola."

TABLE IV

THE AVERAGE PER CENT. OF OBSERVERS CONFUSED BY THE WORD "CHERO-COLA," THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS, WHEN APPLIED TO THE NAMES OF THEIR ARTICLES

Number of Decisions	Decision	Confused Av. P.C.
	Coca-Cola—Chero-Cola	*40.*
5	Infringement	33.2
4	Non-infringement	32.5

If we compute the average percentages of the five infringing marks, and of the four non-infringing marks, as in Table IV, we find that the score of "Coca-Cola—Chero-Cola" is higher than either average. It is about 7 per cent. higher. The average for the infringements is 33.2 per cent., and for the non-infringements 32.5 per cent. The average of the infringements is only 0.7 per cent. higher than the average of the non-infringements.

The most important characteristic of this difference is its small amount. On the other hand, "Green River—Green Ribbon" and "Gold Dust—Gold Drop" is 20 per cent; and the largest difference between two non-infringements, "Sozodont—Kalodont" and "Grape-Nuts—Grain-Hearts" is 30 per cent. Returning for a moment to Table III, two non-infringements, three if we include "Coca-Cola—Chero-Cola," are more confusing than four infringements. The most confusing imitation is a non-infringement; and three of the four most confusing imitations are non-infringements, including "Coca-Cola—Chero-Cola." The least two confusing imitations are non-infringements. The differences within both classes of decisions so enormously outweigh the difference between the classes, that for practical purposes the difference in confusion between the infringing and non-infringing imitations may be disregarded.

Therefore, these decisions are not entirely reliable or consistent. The experiment, presenting conditions no more difficult or essentially different from those in daily life, finds that there is a likelihood of confusion with all non-infringing imitations which the courts held to show no confusion.

d. Duplicate and New Trade-Marks. By calculating how well the duplicate and new trade-marks were recognized, the difficulty of the task set the observer may be shown. If the observer recognizes correctly all the duplicates and new trade-marks, he would then obtain for them 100 per cent. of correct recognitions, or a perfect score. A score of 50 per cent. of correct recognitions may indicate that the experiment is exceedingly difficult or that the observer was simply guessing, in either case meaning zero or no memory of the duplicate shown in the presentation. By pure chance alone it would be possible to obtain this score. An average of 75 per cent. of correct recognitions would then stand midway between zero ability to recognize and just a perfect score; an average of 85 per cent. would stand seven-tenths of the distance to the latter. The other intervening scores can likewise be judged as to their relative difficulty. So, irrespective of the observer's recog-

nitions of the imitations, the difficulty of the experimental conditions may be shown by calculating the per cent. of correct recognitions of the duplicate and new trade-marks.

As is shown in Table V, the duplicates have a score of 80 per cent., and the new trade-marks a score of 94 per cent.

TABLE V

THE AVERAGE PER CENT. OF OBSERVERS THAT CORRECTLY RECOGNIZED THE DUPLICATE AND NEW MARKS, WITH ONE IMITATION WITHOUT THE NAME OF THE ARTICLE

Number of Marks	Kind of Mark	Average Per Cent
19	Duplicate	80
20	New	94
	Average	87

It is thus easier to tell the marks that have not been seen before than those that have. The final average 87 per cent. represents the difficulty of the task set the observers. That is, excluding the results of the imitation, the average per cent. of correct recognitions of all the other marks is 87 per cent. The task was perhaps not more difficult than that which the prevailing conditions in daily life would offer, nor so easy that it could be reacted to without error.

TABLE VI

THE AVERAGE PER CENT. OF OBSERVERS THAT CORRECTLY RECOGNIZED THE DUPLICATE AND NEW MARKS, WITH ONE IMITATION WITH THE NAME OF THE ARTICLE

Number of Marks	Kind of Mark	Average Per Cent.
19	Duplicate	66
20	New	91
	Average	79

Table VI presents the results of the duplicate and new trade-marks in the experiment in which one imitation appeared, and in which the trade-marks were used with the names of their commodities. The duplicates have a score of 66 per cent., and the new trade-marks a score of 91 per cent. In Table VI, as also in the preceding table, the average per cent. of the new trade-marks is higher than that of the duplicates. There is, however, in Table VI a lower score for the duplicates than in Table V; those for the new trade-marks are about the same.

This decrease in per cent. of correct recognitions for the duplicates with the name of the article is explained by the fact that in the same length of exposure more material in the presentation had to be taken in than when the trade-mark alone was shown. A weaker mental impression for the duplicates in Table VI thus caused a drop in the per cent of correct recognitions. But, adding the name of the article to the new trade-marks made just a slightly lower change in ability to discriminate them. We have already seen above in Table II that the relatively shorter time to receive the impression of the original with the name of the commodity aided in causing more confusion than did the original trade-mark without it, as in Table I. The final average in Table VI is 79 per cent. Although it is 8 per cent. lower than the final average in Table V, the task is still of fair difficulty.

TABLE VII

THE AVERAGE PER CENT. OF OBSERVERS THAT CORRECTLY RECOGNIZED THE DUPLICATE AND NEW MARKS, WITH 10 IMITATIONS WITH THE NAME OF THE ARTICLE

Number of Marks	Kind of Mark	Average Per Cent.
10	Duplicate	68
20	New	94
	Average[1]	85

Table VII gives the average per cent. of correct recognitions of the duplicate and new trade-marks in the experiment in which 10 imitations appeared, and in which the trade-marks were used with the names of their commodity. The duplicates have a score of 68 per cent., and the new trade-marks a score of 94 per cent. As in Tables V and VI the average of the new trade-marks is higher than the average of the duplicates. The scores in Tables VI and VII are for both kinds of marks about the same. The final average 85 per cent. in Table VII represents the difficulty of the task.

B. SOUND CONFUSION MEASURED BY RELATIVE POSITION

1. *General Procedure and Technique.* The method of relative position was used to compare the likelihood of confusion in sound between the words "Coca-Cola" and "Chero-Cola" with that between the litigated trade-marks of the previous

[1] This average was obtained by dividing the sum of the 30 separate scores by 30; it is slightly higher than the average obtained by dividing the sum of the average of the duplicate and new scores by 2.

experiment. This method does not state "how many" are confused in sound by the various imitations, but it does give a measure of their relative differences in sound confusion. It required the observer to arrange the 10 pairs of litigated trade-marks in a list according to the amount or magnitude of confusion that the imitation is likely to cause. Any pair is then measured by its position in the list. In the comparison of "Chero-Cola" with the nine imitations, and the psychological criticism of their legal decisions it is recognized that the decisions were not rendered entirely on sound confusion.

Measurement by relative position was adopted: first, because it is a different method from that used in the previous experiment; second, because there is a certain analogy between the mental processes of the observer judging relative differences and the mental processes in the minds of the court judging the question of infringement or non-infringement; and third, because confusion between the words "Coca-Cola" and "Chero-Cola" is especially likely to occur under certain actual business conditions in respect to their sounds.

As regards the first point, it would obviously be of much importance if a method different from that employed in visual recognitive confusion should show that confusion in sound between the words "Coca-Cola" and "Chero-Cola" was above that of the average of the infringing and non-infringing marks. Confirmation of greater likelihood of confusion than infringements and non-infringements by two different methods is certainly significant.

As regards the second point, the mental processes of the observer comparing relative differences in sound confusion are essentially very similar to those in the minds of judges comparing the sound confusion of a pair of trade-marks in a case in court with that of trade-marks adjudicated infringement and non-infringement. Although there are resemblances between the mental conditions that give the experimental data and those that give the legal decisions, there are differences between this experimental method and the legal procedure that have greater significance. With measurement by relative position the accuracy or mathematical validity is higher. In employing a far greater number of observers the arrangement is much less likely to be the result of chance bias of a few individuals. In the average of a large number of judgments a chance bias in any one direction is cancelled, and the result represents the general tendency of all the observers.

Furthermore, the observers are entirely disinterested in the outcome of the experiment. They do not know which trade-

marks have been adjudicated infringement or non-infringement, nor that any pair of trade-marks is being especially investigated. Not knowing that the experiment is being conducted for a pending case the judgments of the observers are consequently without prejudice to either plaintiff or defendant. On the other hand, with legal procedure not more than a few judges pass on the decision; and their judgments tend to some extent to be influenced by the abilities of contending counsels to magnify the differences and increase the similarities of the trade-marks. Measurement by relative position gives an exact measure, whereas a judicial decision throws a case into one or two categories, ill-defined and without quantitative significance. It is thus clear that this experimental method is far superior to the present legal procedure.

As regards the last point, the probability of confusion in sound or pronunciation between the words "Coca-Cola" and "Chero-Cola" is considered because it is in this respect that a purchaser asking for a glass of "Coca-Cola" from the dispenser of the drink at the fountain is likely to be fooled. The purchaser under these conditions does not usually see the label on the bottle or its crown, and and he can not identify the drink by its trade-mark. Assuming that the dispenser is acting honestly in listening to his customer's request, and in fulfilling what he believes to be that request, confusion between the words "Coca-Cola" and "Chero-Cola" is most likely in sound or pronunciation alone.

Each observer was given an envelope and a sheet of directions. Enclosed in the envelope were 10 slips of white paper, on each of which appeared two trade-marks. The slips of paper were the same size as those of the previous experiments; and the trade-marks were all typewritten in black ink. The name of the article or commodity did not appear with the trade-mark. The following reproduction is a sample of one of the slips and is identical with that actually used.

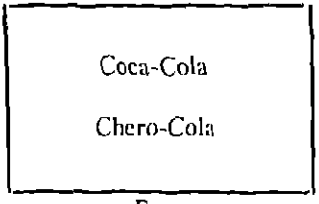


FIG. 2

The directions were as follows:

"Please arrange in serial order as well as you can the enclosed slips of paper according to the following instructions.

"Each slip contains two trade-marks of a common article. The upper is the original trade-mark, and the lower is the imitation of it.

"The different pairs of trade-marks on the different slips vary in their likelihood of confusion, or in their deceptive similarity, or in the likelihood that the imitation trade-mark will be mistaken for the original. *This confusion may be due to the appearance of the trade-marks, their sound, their linguistic formation, their meaning or significance, or any combination of these four factors.* Neglect, however, all of these four factors except that of *sound*.

"Arrange the slips according to the likelihood of confusion in just the *sound* of the trade-marks: Put on the top of the pile the slip containing the imitation which shows the greatest likelihood of confusion in *sound* and on the bottom the slip containing the imitation which shows the least likelihood of confusion in sound. Between the top slip and the bottom put in order those slips containing imitations which show intervening degrees of likelihood of confusion in sound.

"In arranging the slips consider the trade-marks as if you have had no previous experience with them."

In addition, the observer was told that there was no time limit, and that he could take as long as he wished to make the arrangement, and further that he was allowed to rearrange the order until it satisfied him. To do the experiment usually required about five minutes. No information was given as to the purpose of the experiment.

Fifty observers assisted in the task of judging the trade-marks. They were mostly students of psychology at Columbia University, during the month of October, 1915. Some were undergraduates studying psychology, and some were graduate students with special training in it, and a few were assistants and instructors of psychology. Only four observers had not gone to college, but all had been acquainted with psychological experiments. No observer that took part in any of the recognition experiments acted again in this experiment.

2. *Results.* Table VIII gives the grades and probable errors of the word "Chero-Cola," each of the five infringing, and four non-infringing imitations. The first column in the table shows whether the imitation in the decision was held to be an infringement (I) or a noninfringement (N). The second column gives the original trade-mark in the decision; and

the third the imitative. The fourth column gives the grade; and the fifth and last the probable error of the grade. The marks are arranged in an order of probable confusion from most to least. The pair of trade-marks which received the smallest average was judged to have the greatest likelihood of confusion in sound; the pair which obtained the largest average was judged to have the least likelihood of confusion in sound.

TABLE VIII

THE GRADES AND PROBABLE ERRORS OF THE WORD "CHERO-COLA" EACH OF THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS

Decision	Trade-Mark Original	Trade-Mark Imitative	Av.	P. E.
I	Cascarets	Castorets	1.2	.03
I	Green River	Green Ribbon	3.3	.13
	Coca-Cola	*Chero-Cola*	*3.6*	*.17*
I	Listerine	Listogen	3.7	.15
N	Sozodont	Kalodont	5.1	.20
I	Gold Dust	Gold Drop	5.2	.12
N	Grape-Nuts	Grain-Hearts	7.0	.13
I	Club	Chancellor Club	7.8	.11
N	Mother's	Grand-Ma's	8.6	.09
N	Holeproof	Knotair	9.5	.08

"Cascarets—Castorets," an infringement, has been selected out of the 10 pairs of trade-marks as being the most confusing. It was placed in the first position by 42 observers, and in the second by 8. Its average grade or position is 1.2. "Green River—Green Ribbon" stands second with a grade of 3.3. "Coca-Cola—Chero-Cola" with a grade of 3.6 is nearly tied by an infringement, "Listerine—Listogen" with a grade of 3.7 for the third most confusing position. The grade of "Coca-Cola—Chero-Cola" is 0.3 lower than that of "Green River—Green Ribbon," and 0.1 higher than that of "Listerine—Listogen." These differences are not large. The grade of "Coca-Cola—Chero-Cola" is 2.4 lower than the grade of the most confusing infringement "Cascarets—Castorets;" and it is 4.2 higher than that of the least confusing infringement "Club—Chancellor Club." It is 1.5 higher than the most confusing non-infringement "Sozodont—Kalodont;" and it is 5.9 higher than the grade of the least confusing non-infringement "Holeproof—Knotair." The arrangement shows that "Coca-Cola—Chero-Cola" is about equally confusing with one infringement, less confusing than two infringements, and more confusing than two infringements and all four non-infringements.

A consideration of the probable errors will show that the grades of confusion of the different pairs of trade-marks have a high degree of validity. Moreover, the validity itself of the order can be measured. The probable error of the grade 1.2 is 0.03. It must, however, be remembered that the chances of the true grade being far outside this range decrease very rapidly. The possibility that the grade 1.2 of "Cascarets—*Castorets*" is due to chance is entirely negligible. Although all the remaining probable errors are somewhat larger, they indicate a high degree of validity of the average grades. The larger probable errors indicate that the order is less certain, and the difference in confusion between one pair of trade-marks and the next on the list is less. The probable error of the grade 3.6 of "Coca-Cola—Chero-Cola" is 0.17, or the chances are even that this grade is correct within about one-fifth of a unit.

The probable error of the difference 0.3 between the grade 3.6 of "Coca-Cola—Chero-Cola" and the next higher grade of "Green River—Green Ribbon" will show the chances that "Coca-Cola—Chero-Cola" deserves a grade as high as that of "Green River—Green Ribbon." The probable error of the difference 0.3 is 0.21. As the difference 0.3 is about 1.4 times greater than the probable error of the difference 0.21, the chances are about 3 to 2 that "Coca-Cola—Chero-Cola" does not deserve a grade as high as that of "Green River—Green Ribbon." As the difference 0.1 between the grade of "Coca-Cola—Chero-Cola" and the next lower grade of "Listerine—Listogen" is covered by the probable errors of the grades, the chances are nearly even that "Listerine—Listogen" deserves a grade as high as that of "Coca-Cola—Chero-Cola." The probable error of the difference 1.5 between the grade of "Coca-Cola—Chero-Cola" and the second lower grade of "Sozodont—Kalodont" is 0.26. As the difference 1.5 is about 5.8 times greater than the probable error of the difference 0.26, the chances are over 10,000 to 1 that "Coca-Cola—Chero-Cola" does not deserve a grade as low as that of "Sozodont—Kalodont." Therefore, the chances are about 3 to 2 that "Coca-Cola—Chero-Cola" does not deserve a grade as high as that of "Green River—Green Ribbon;" the chances are nearly even that "Listerine—Listogen" deserves a grade as high as that of "Coca-Cola—Chero-Cola;" and the chances are over 10,000 to 1 that "Coca-Cola—Chero-Cola" does not deserve a grade as low as that of "Sozodont—Kalodont."

Of the very well-known original trade-marks "Cascarets—

Castorets" alone stands higher than "Coca-Cola—Chero-Cola," and "Gold Dust—Gold Drop" and "Grape-Nuts—Grain-Hearts" are both lower. Of the less well-known original trade-marks "Green River—Green Ribbon" stands higher than "Coca-Cola—Chero-Cola," "Listerine—Listogen" about the same, and all the rest lower. In this relative position experiment the comparison between these two classes of trade-marks is not really necessary as the instructions required the observer to disregard the factor of familiarity with the trade-marks. In only three cases did the observer report any difficulty in so doing it.

TABLE IX

THE AVERAGE GRADES AND PROBABLE ERRORS OF THE WORD "CHERO-COLA," THE FIVE INFRINGEMENTS, AND FOUR NON-INFRINGEMENTS

Number of Decisions	Decision	Av.	P. E.
	Coca-Cola—Chero-Cola	3.6	.17
5	Infringement	4.2	.12
4	Non-infringement	7.6	.11

Table IX gives the average grades of "Coca-Cola—Chero-Cola" of the five infringements, and four non-infringements. "Coca-Cola—Chero Cola" has a higher rank than both the average grades of the infringements, and the average grades of the non-infringements. It is 0.6 higher than the average of the infringements, and 4.0 higher than the average of the non-infringements. The average of the infringements is 4.2, and that of the non-infringements 7.6; their probable errors are 0.12 and 0.11 respectively.

The probable error of the difference 0.6 between the grade of "Coca-Cola—Chero-Cola" and the average of the five infringements is 0.21. As the difference 0.6 is about 2.9 times greater than the probable error of the difference 0.21, the chances are about 19 to 1 that "Coca-Cola—Chero-Cola" does not deserve a grade as low as the average of the five infringements. The probable error of the difference 4.0 between the grade of "Coca-Cola—Chero-Cola" and the average of the four non-infringements is 0.20. As the difference 4.0 is 20 times greater than the probable error of the 0.20, there are practically no chances that "Coca-Cola—Chero-Cola" deserves a grade as low as the average of the four non-infringements. "Coca-Cola—Chero-Cola," therefore, stands with high validity above the average of the five infringements, and the average of the four non-infringements.

In Table VIII, among the infringements the greatest difference is 6.6, between "Cascarets——Castorets" with a grade of 1.2 and "Club—Chancellor Club" with a grade of 7.8. Among the non-infringements the greatest difference is 4.4, between "Sozodont—Kalodont" with a grade of 5.1 and "Holeproof—Knotair" with a grade of 9.5. In Table IX the average of the five infringements is 3.4 higher than that of the four non-infringements. Thus the differences within either class of decisions are greater than between them. Two non-infringements, three if we include "Coca-Cola—Chero-Cola," are more confusing than the lowest infringement. The two most confusing imitations are infringements, and the two least confusing non-infringements. In sound confusion the results point consequently as did those in visual recognitive confusion to the unreliability and inconsistency of the legal decisions and of the present legal procedure. The difference, however, between the averages of the infringements and non-infringements in sound confusion by relative position is greater than in visual recognition confusion, and indicates a greater accuracy of the decisions.

A comparison of Tables III and VIII will show that the positions of the same pairs of marks in the two lists have changed in most cases. "Green River—Green Ribbon" is the only pair that did not change its position. "Coca-Cola—Chero-Cola" changed only one-half of a position, "Holeproof—Knotair" changed one position, "Listerine—Listogen" and "Gold Dust—Gold Drop" two positions, "Grape-Nuts—Grain-Hearts" and "Club—Chancellor Club" three, "Sozodont—Kalodont" four, "Mother's—Grand-Ma's" five and one-half, and "Cascarets—Castorets" six. The average of the number of positions changed for the 10 pairs of marks is 2.7. The lack of exact correspondence between the positions of the same pairs in both lists does not mean that the methods or results are contradictory or inconsistent. Though in both experiments presentation of material was through vision, each experiment principally affected two different senses. Exact correspondence would most likely have existed between the two lists if they were both concerned with measuring the same thing.

Conclusion

1. The word "Chero-Cola" shows a likelihood of visual confusion with the word "Coca-Cola." With one imitation in the experiment and when the mark is used without the name of the commodity, 28 per cent. of all the observers confuse in visual recognition the word "Chero-Cola" with "Coca-

Cola." With one imitation in the experiment and when the mark is used with the name of the commodity, 68 per cent. of all the observers confuse in visual recognition the words "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink." With 10 imitations in the experiments and when the mark is used with the name of the commodity, 40 per cent. of all the observers confuse in visual recognition the words "Chero-Cola Soft Drink" with "Coca-Cola Soft Drink." Thus, either with or without the name of the commodity added to the trade-mark, with or without other imitations present, the word "Chero-Cola" shows a likelihood of confusion in visual recognition with "Coca-Cola." It should be observed that the per cent. of likelihood of confusion is not an absolute number representing the exact per cent. of individuals that would be confused in daily life. The per cent. holds approximately under the simple standardized conditions of the experiment.

2. Most of the observers confused are "absolutely certain" that they had seen the word "Chero-Cola," and only a few are either "reasonably certain" or have a "faint idea." That is, most of the observers put the greatest confidence in their wrong recognitions, and in so doing testify to or affirm their great belief that they are right.

3. Most of the observers confused state that they were "perfectly familiar" with the trade-mark "Coca-Cola" before they saw it in the presentation of the experiment, and *only a few were either "moderately familiar" or "just familiar;"* none were "unfamiliar" with it.

4. Confusion in visual recognition caused by "Chero-Cola" is greater than that caused by the imitations of the other three very well-known original marks; it is 12 per cent. greater than the confusion caused by the infringing imitation of "Cascarets," 14 per cent. greater than that of the infringing imitation of "Gold Dust," and 22 per cent. greater than that of the non-infringing imitation of "Grape-Nuts." "Chero-Cola" has a score of 8 per cent. less than the most confusing non-infringement, 6 per cent. less than the most confusing infringement, ties with another non-infringement, and has a higher score than four infringements and two non-infringements. The confusion in the case of "Coca-Cola—Chero-Cola" is about 7 per cent. greater than that of the average of the five infringing pairs of marks or of the four-non-infringing. The average of the infringements is only 0.7 per cent. higher than that of the non-infringements; practically this is no difference.

5. As determined by the average per cent. of observers that correctly recognized the duplicate and new marks, the task set the observer in the experiment is of moderate difficulty. In the three recognition experiments the average per cents. of correct recognitions of the duplicate and new marks is 87, 79, and 85 per cents. The figures indicate then that the task set is not too arduous, nor exceedingly easy.

6. In measurement of confusion in sound by relative position "Coca-Cola—Chero-Cola" with a grade of 3.6 (P. E. 0.17) is in the third most confusing position in the list of 10 pairs of litigated trade-marks. It is ranked nearly equally confusing with one infringement, less confusing than two infringements, and more confusing than two infringements and all four non-infringements.

7. "Coca-Cola—Chero-Cola" has a higher rank of confusion than both the average of the grades of the infringements, and the average of the grades of the non-infringements. It is with high validity above the average 4.2 (P. E. 0.12) of the five infringements, and above the average 7.6 (P. E. 0.11) of the four non-infringements.

THE ANALYSIS OF TRADE ABILITY

By Edward S. Robinson, Yale University

I. *Introductory*

The control of learning and the measurement of acquired abilities are among the most important personnel problems before the industrial psychologist. He will find that these problems hold fully as great opportunities for the application of the method of his science as do the problems of intelligence testing. Although he has no ready-made formulas for dealing with the situation, he has a psychological, or better, an empirical point of view toward human behavior. By virtue of this point of view he can, if he will, contribute a systematic analysis of many problems of industrial behavior which seem so stubbornly to resist analysis on the part of the average man who is familiar with industry, but unfamiliar with psychological methods.

The possibility of applying psychological methods to the problems of acquired behavior in industry can best be illustrated by reference to the work of the late Trade Test Division of the Committee on Classification of Personnel in the Army. This organization had as its task the development of standardized tests for measuring the trade ability of men entering the National Army. It was the purpose of this work to increase the accuracy of selecting men best fitted for particular army jobs requiring special abilities. Civilian occupations seldom furnished exact counterparts of these army jobs, but correspondence was close enough and frequent enough to give a very high value to any information which could be obtained regarding a man's industrial experience. The trade tests were designed to check up on statements of soldiers concerning their pre-military occupations, and, further, to measure their proficiency in trades at which they claimed to have worked. The development of these tests required the observation, from a psychological point of view, of a great number of trades and tradesmen. I shall try to summarize here the results of those observations in so far as they apply, not to specific trades, but to trades in general.

Although it has some basis in a rather loose usage, the meaning of the term *trade* as used in this discussion is largely arbi-

trary. It does not, of course, refer to occupations mainly concerned with the exchange of goods, but rather to occupations emphasizing an acquired facility in the use of certain tools, instruments and machines for the achievement of certain classes of physical results. Moreover, this acquired facility or skill consists of a complex set of coordinations, intellectual, motor, or both, such as we find in carpentry, interior wiring and the like, and not of the very simple and oft repeated coordinations typical of the operation of much standardized factory machinery. *Trade ability* refers, then, to any complex set of coordinations which we find characterizing all men skilled in a given trade and marking them off as a single group.

II. *Defining Trades*

There are in existence scores of books and pamphlets which attempt to define the trades of modern industry, but scientifically most of these are of little or no value. Practically all of the writers of these definitions have fallen into an easy fallacy which appears when we look into their methods. Having a given title, say armature winder, before them, they have set out to collect from all possible sources descriptions of the jobs of armature winders. They have then tabulated these descriptions and issued them as a full and complete definition of the armature winder—as a guide for identifying the skilled armature winder wherever he may be found. But when we try to make such practical use of the definition, we find that it does not work. Only in negligible instances do we find the type of man we are looking for. According to the definition a skilled armature winder is a man capable of working on "A.C." and "D.C." windings, on "hand" and "form" windings, etc. Our natural inference from this is that the man who only partially fills these requirements is only partially skilled. We find, however, that there are men with all sorts of partial combinations of these enumerated requirements who are, nevertheless, classed as winders of the first rank by their bosses and co-workers. From this it is apparent that "armature winding" is not a single trade, but a name which has been applied indiscriminately to a variety of trades. The trade of armature winding, pure and simple, does not exist.

In order to frame useful trade definitions, then, the investigator must be thoroughly empirical. There is only one reliable source of information about the requirements of a trade and that source is to be found in the observed abilities

of individual tradesmen. Industrial experts of one kind or another may safely be consulted for a bird's-eye view of a trade whose definition is required, but such experts are not to be relied upon for detailed or exact opinions. The classification and description of workers according to their peculiar behavior equipments is a thing which must wait upon the field observations of men who have the training and patience to be inductive and empirical. Such a procedure will result in definitions or specifications quite different from those to be found in the large mass of current trade literature. We shall find that the tradesmen who can actually be identified in industry are "A.C." men, "D.C." men, form winders, hand winders, or general repairmen, etc. and not armature winders.

III. *Elements of Trade Ability*

The detailed work of tabulating the pertinent items of behavior of individual tradesmen requires a definite basis of analysis. In the Trade Test Division we found it convenient to conceive of ability in a trade as made up of a group of elementary abilities. These simple abilities, while not necessarily unanalyzable, are of such a nature that when further reduced they bear no apparent relationship to the complex trade ability of which they are parts. The riveting of the sheet metal worker is an element of his trade ability, although the behavior here involved is evidently open to further analysis. Such further analysis would, however, yield items of behavior with no apparent relationship to sheet metal working ability and, therefore, of irrelevant fineness. In all trades analysis stopped with the discovery of the simplest relevant units of ability.

From a psychological point of view the analysis of any trade brings out elements of two main sorts: (1) *Intellectual*, and (2) *Motor*. While in the strictest theoretical sense no sharp line can be drawn between these two categories, we may say roughly that the intellectual elements are the items of information and habits of judgment employed in meeting trade situations. Such items of information are displayed when a carpenter recognizes a mitre, a shingle hatchet, a valley rafter etc. Typical habits of judgment are displayed when an interior wireman selects one rather than another rosette as a ceiling fixture for a drop cord and when he decides to cover a bit of wire with porcelain tubing instead of lume. The motor elements are the coordinations involved in the unique muscular movements of tradesmen. The wood pattern maker possesses a large variety of nice coordinations which are

manifest in his manipulation of numerous chisels, planes and other tools. Shifting geers, accurate steering, etc., are motor elements of automobile driving.

When the Trade Test Division first faced the problem of measuring trade ability, a very confident opinion was expressed in all quarters as to the relative importance of intellectual and motor elements in trades in general. Psychologists, engineers, and tradesmen themselves insisted that the skilled tradesman is characterized by his manual rather than by his intellectual training, and that, on the other hand, there are large numbers of men having a thorough intellectual grasp of trades they cannot practice. The point seemed well taken and influenced appreciably the direction of our earlier attacks upon the problem of measurement. Concretely, this point of view appeared in an a priori confidence that performance tests would measure trade ability more accurately than oral or other tests emphasizing information and judgment. But increasing knowledge, gained from first hand observations of industrial conditions, soon made clear that this weighting of motor elements was based upon an outgrown state of affairs. Undoubtedly it was once true that ability in perhaps a majority of skilled trades was characterized by motor habits, to be acquired only after long and arduous training. This, however, is no longer the case. Pattern making, automobile driving and some other present day trades emphasize the motor, but far more common are the electrical trades, the machine trades, etc., which are largely intellectual and whose motor habits, with few exceptions, are relatively simple and easily acquired.

This statement that trades are becoming more intellectual may need reconciliation with the fact that modern industry calls for little intelligence from the run of her workers. Especially since the beginning of the war manufacturers have reduced the complexity of the duties of individual workers by cutting up trades into operations. Under such conditions it has been possible to use practically unskilled labor in the production of intricate mechanical products. But where jobs have remained complex and where workers have remained tradesmen in the sense in which I have defined that term, there has been a growing emphasis upon the intellectual as opposed to the motor. Each new machine introduced into industry tends further to emphasize the knowledge and judgment requirements of workers and to do away with skilled movements previously required.

IV. *Variations in Trade Ability*

In any trade ability ranges from that of the novice to that of the expert. The way in which trade ability develops can conveniently be studied by listing the elementary abilities manifest in men whose industrial status indicates just what stage they have reached in the learning process. In the novice we find no item of knowledge, no habit of judgment, no muscular movement which is a genuine element of the trade in question. The presence of such an element in a novice indicates one of two things: either the novice is not a novice, but an immature tradesman; or this particular element is extraneous to the trade and should, therefore, be excluded as a partial criterion of ability in that trade. In the expert we find a maximum number of genuine trade elements. In men whose status is something between novice and expert we find a partially complete group of trade elements.

The building up of a group of trade elements—the learning of a trade—takes place in a definite, fairly predictable way. The order in which the individual elements are acquired is not a matter of chance, but for the most part is the same for all tradesmen within a single trade. That is to say, most trade elements tend to appear at some definite stage in the learning process. Hardly any apprentice battery mechanic can do a successful job of "lead burning" on the top of a battery, while this same task is easy for almost any battery mechanic who has passed the apprentice stage. Similarly, technical knowledge of roofing appears somewhat suddenly in the process of learning house carpentry. These elements which tend to appear suddenly may be classified in the following fashion:

1. Trade Elements: Those possessed by practically all workers in a trade—apprentices, journeymen and experts—and by no one not in that trade.

2. Journeyman Elements: Those possessed by practically all journeymen and experts and by no apprentices or novices.

3. Expert Elements: Those possessed only by experts.

If for an element of any of these types we should plot the correlation between the probability of its appearance and trade status, we would obtain a sharply curved regression. Figures 1, 2 and 3 show the plotted correlations for hypothetical elements of each of the above types.

In our studies we found that the most frequent elements were those belonging to the first of the above classes. Elements tending to appear at the journeyman stage were next most frequent. Elements tending to appear at the expert

stage were quite rare. Increments of trade elements, then, grow smaller and smaller as progress is made toward expertness, and trade learning as measured by the acquisition of new elements is a negatively accelerated function.

Besides these elements, each of which has a definite point of appearance in the progress of learning, there are other elements whose probability of appearance, when correlated with trade status, gives a linear regression. Figure 4 shows a hypothetical element of this type. Such elements can be described as appearing in no novices, in a few apprentices, in more journeymen, and in still more experts. In our work we ran across comparatively few elements of this kind.

V. *The Measurement of Trade Ability*

Having once classified the elements of a given trade the construction of a test for that trade is simple. If the object of the test be that of separating men on either side of a point along the range of ability, it is necessary to include in that test only questions or job specifications calling for a display of elements tending to appear at just beyond that point. If, on the other hand, the object be that of separating men at several points along the range of ability, then the test should include elements tending to appear at just beyond each of these points, or elements whose probability of appearance increases gradually along the range of ability, or elements of both classes.

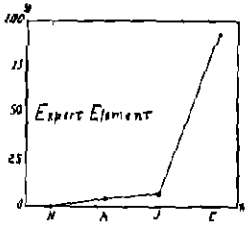


Fig. 3

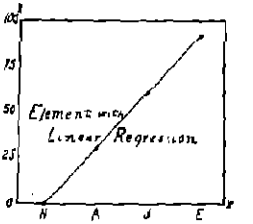


Fig. 4

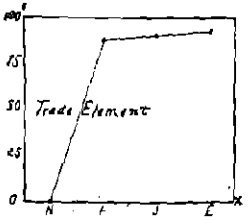


Fig. 1



Fig. 2

x axes show stages in trade learning: N = novice, A = apprentice, J = journeyman, E = expert

y axes show probability of appearance in terms of percent

A WRITTEN TRADE TEST: MULTIPLE CHOICE METHOD

By J. Crosby Chapman, Western Reserve University; and Herbert A. Toops, Teachers College, Columbia University

The Army Trade Test Division was, on *a priori* grounds, extremely sceptical of written trade examinations. For this reason, in the theoretical examination, an oral test was used. The oral method has proved its power to distinguish, with a fairly high degree of certainty, between novices, apprentices, and journeymen, and to a less degree between journeymen and so-called experts. The possibilities and limitations of one type of written test form the subject of this article. In connection with the work of this Division an opportunity was given of trying out a written trade examination constructed on the multiple-choice plan. There are many advantages in being able to give group examinations of this kind. Among these are: (1) conservation of examiner's time, (2) reduction of scoring time and increase of scoring accuracy through the use of stencils, etc., (3) elimination of subjective judgments, (4) the possibility of rapidly eliminating the trade "bluffer," leaving only those who show evidence of some knowledge for further consideration or examination by means of oral, performance, or picture tests.

Assembly of Questions

In conference with expert tradesmen, sixty-two questions were prepared, each question having four possible answers, of which one was correct. Nearly all the questions of the army oral test for Bricklayers, partly collected by the authors, were adapted to this form of examination, and the remainder were secured from other sources. Trade terms were used almost exclusively in the four answers from which the choice was to be made, on the supposition that a person might be able to recognize a trade term from among a number of terms not applicable to the trade and yet have no real knowledge of its significance. A part of the test is reproduced.

BRICKLAYERS' TEST*

Directions:

In the sentences below, there are four choices for a correct sentence. Only one of the four choices is correct. Draw a line under the *one* choice which makes the best sentence. If you are not sure, *guess*; an omitted answer will count as a wrong answer in determining your score. The first two sentences, A and B, are answered correctly as a sample for you. Read them carefully, then go on to question 1, question 2, and so on to the end of the list.

Samples:

A. A half of a brick is called a: chunk, block, heel, bat.

B. Fire-bricks are laid in: concrete, cement, fire-clay, mortar.

Questions:

1. The top course of stone on a wall is called: coping, bond-stone, clip-course, capstone.
2. A brick set on end is called: upright, soldier, row-lock, stud.
3. Before plumbing up a corner, you should lay: 3-courses, 6-courses, 9-courses, 12-courses.
4. A brick that is set on the narrow edge is called: stretcher, oarlock, row-lock, header.
5. In coming to a height, if there is a course of brick difference in the level, you would call it a: haunch, filler, line-level, hog.
6. Filling-in the space between a front and back course of brick is called: slushing-up, tempering, tuck-pointing, plastering.
7. Bricks laid across the wall reaching to the face are called: stretchers, headers, layers, fillers.
8. A fire stop around a flue is formed by a: coping, skew-back, corble, indent.
9. If you have no metal ties in plain bond work, the front work of a building is tied-in with: king-course, clip-course, projection-course, belt.
10. In backing a 4-inch ashlar stone wall, you would tie it with: tie-joint, binder, brace, anchor.
11. A brick cut in half lengthwise is called: filler, soap, bat, chunk.
12. To keep the line level in the middle of a long wall you use: level, plumb-line, square, trigger.
13. In cutting a brick you would use a: chisel-bar, striker, brick-set, cutter.

---

*These eighteen questions were selected from the original set of sixty-two, which constituted the examination. This selection was made on the basis described in the text.

14.	The distance above the frame for a lintel should be:	½-inch, ¾-inch, 1-inch, 1½-inches.
15.	An ordinary stone sill should be bedded on the:	front, back, solid, two ends.
16.	To locate a building you use:	water-level, sighter, transit, line-level.
17.	A skutch is used in cutting:	cement, glazed-brick, stone, fire-brick.
18.	Another name for a blind header course is:	belt-course, clip-course, dental-course, hidden-course.

The answers to the above questions with the percentages of experts passing each are as follows:

1.	Coping—98%	7.	Headers—98%	13.	Brick-set—100%
2.	Soldier—98%	8.	Corble—85%	14.	½-inch—88 %
3.	3-courses—88%	9.	Clip-course—100%	15.	Two-ends—93%
4.	Rowlock—100%	10.	Anchor—85%	16.	Transit—88%
5.	Hog—100%	11.	Soap—85%	17.	Fire-brick—100%
6.	Slushing-up—98%	12.	Trigger—93%	18.	Clip-course—93%

The Subjects of the Experiment

The examination was given without time limit to 101 persons, distributed as follows: 35 novices without trade experience, 4 apprentices, and 62 trade union members. Of the novices, 29 were soldiers in an army training school, being instructed in other trades at the time, but inexperienced in the bricklayers' trade. These men were mechanically inclined, and above average in intelligence and education. It is obvious, therefore, that the questions were subjected to much more intelligent guessing than ordinary conditions, under which they would be employed, would furnish. Of the other six novices, 3 were stenographers who had several days experience, immediately preceding this test, in taking and transcribing stenographic notes relating to the general subject of bricklaying. The other 3 novices were clerical workers in the Department of Psychology of Carnegie Institute of Technology. These subjects were well informed concerning the technique employed in assembling such tests, and used various methods of "suggestion, inspection, and elimination" in efforts to make as high scores as possible.

Owing to the great scarcity of apprentices, only four could be examined. Their length of trade experience ranged from one month to five years. In view of experience and performance in the army oral test, the lowest apprentice can be con-

sidered little better than novice, while the best apprentice perhaps deserves to be ranked as a low journeyman.

The tradesmen, being all members of the union, more than fulfil the minimum entrance requirements in point of years of trade experience and skill. On the basis of their own written statements in regard to years of trade experience, type of work, and foremanship ability, these 62 men were divided by the authors into the two following groups: (a) 36 journeymen (hereafter called J.) or men of medium amount of skill and knowledge; (b) 26 experts (hereafter called E.), men of superior skill, knowledge and foremanship ability. This rather arbitrary classification was made in an effort to determine from the results of the examination whether a set of questions could be so selected as to differentiate not only non-tradesmen from tradesmen, and apprentices from skilled men, but also journeymen from journeymen experts, so divided.

Presentation of Results

Four points were allowed for each correct answer; there were no partial credits. Four points was selected as an arbitrary value because of army custom. A graph for each question was drawn, showing the average score made by novices, apprentices, journeymen and experts. Samples of such graphs are shown in fig. 1. Since little was known as to the most desirable combination of questions to make up the best set for an examination, five methods of combining varying numbers of questions were tried; only two of these methods are here described.

Method I is the entire set of 62 questions.

Method II is a selection, from the original 62, of 18 questions, chosen on the basis of the combined judgments of three persons experienced in similiar selection for the army oral trade tests. The requirements for a question in the army tests were: (1) increasing average scores from class to class, novices to experts; (2) a fairly low average novice score; (3) a comparatively high average expert score. The ideal type of question for the choice-method is shown by the straight line of fig. 1.

The distribution of scores by Method I are given in Graph I. By inspection, arbitrary lines of demarcation of N-A, and A-J may be made, corresponding to the *number of questions* correct as follows:

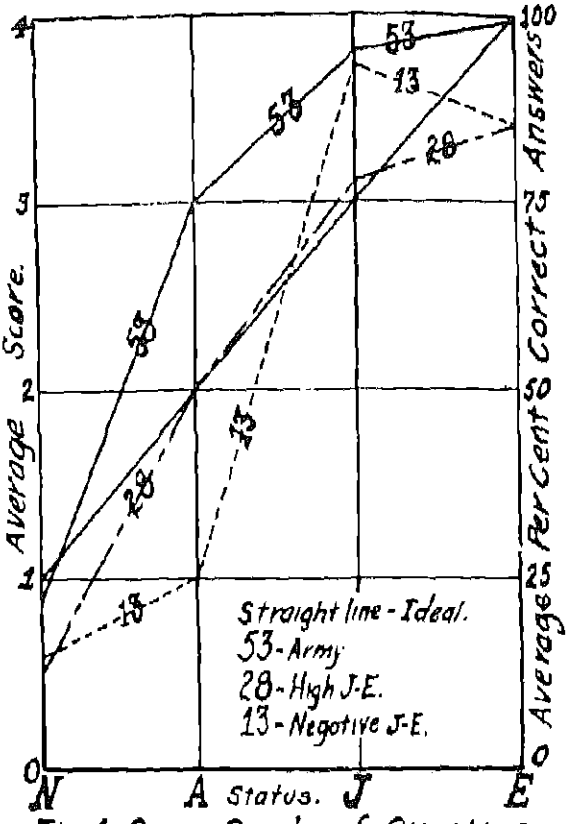
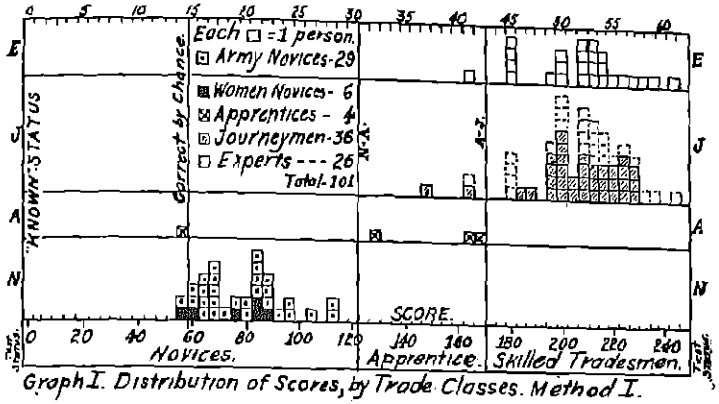


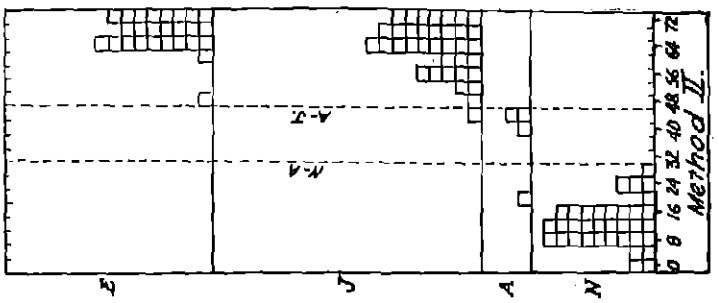
Fig. 1. Some Graphs of Questions.

Novice score	0-30
Apprentice score	31-42
Skilled tradesmen	43-62

It can be seen at once that the test does not differentiate between the journeymen and expert classes. Either the division into these two classes by judgments, based on the men's written statements, has nothing to do with the men's ability to answer this form of question, or the test as given is not sufficiently difficult to differentiate among skilled tradesmen. The homogeneity of the skilled tradesmen class as to scores can be easily seen from Graph I, where the expert distribu-



Graph I. Distribution of Scores, by Trade Classes. Method I.



tion has been added to the journeyman distribution by means of the dotted squares.

With the two arbitrary critical lines just mentioned, only three skilled union men out of 62, or less than 5 per cent, are displaced out of this larger class into the apprentice class. We can assume that all union men have had four years trade experience and possess a certain minimum requirement of ability. The apprentice, with only one month's experience, is here rated with the novices, as is only fair under the conditions. His guessing ability is as poor as the poorest novice, from which we can assume that he has learned next to no trade terms in one month of trade experience. Our own judgment is to the effect that intelligence is a better single measure of trade skill than are years of trade experience, after a certain minimum time for learning has elapsed.

By chance alone, the poorest novice should theoretically be able to answer 15 of the 62 questions correctly (one out of every four). Here, only two novices fail to come up to this standard, and they fail by only a single question each, both answering 14 questions correctly. The average novice answers a little better than 19 questions.

If the object is merely to separate skilled tradesmen from the apprentices, or the tradesmen of experience from the "trade bluffers," the complete set of questions of Method I is a remarkably accurate instrument for so doing. For this purpose, with the persons tested, it is at least 95 per cent efficient.

Method II, which is the selection of the best 18 questions according to the usual army criteria, gives us, with the arbitrary lines of demaracation adopted in Graph II, a slightly better separation of the three classes—novices, apprentices, and skilled union workmen. The one apprentice, with one month of trade experience, is still placed with the novices, but is now better than the average novice, and only one skilled union tradesman (a journeyman) is rated as an apprentice. To make the separations indicated, the test is 98 per cent efficient. This would probably be lowered if we had a fair sampling of apprentices.

Nevertheless, in determining the efficiency of a test for industrial purposes, it seems desirable to weigh the classes according to their frequency of occurrence in industry; if there is only one apprentice for every 25 skilled men in industry, an employment manager would be unable to make many mistakes by hiring apprentices under the impression that they were skilled men.

In industry, also, it seems desirable to measure the efficiency of test methods by taking into account the amount of damage caused by misplacement of a man by the test, some such measure as "dollar displacements," as suggested by R. W. Kelly. For it is a greater industrial mistake to give a journeyman or expert rating to an apprentice than it is to rate a novice as an apprentice; this may be seen from a consideration of the relative amounts of responsibility and wages involved.

From Graph II, Method II, we note that there is a slightly better differentiation of J. and E. than in Method I, but not sufficient to justify any arbitrary line of division between the two classes. If a line were drawn at 60-64, only two experts would be classed as J. while over half of the J. would be classed as E.

The final selection of questions, used in Method II and percentages of experts answering each correctly, are given in the

preceding sample. The average scores and average deviations of the trade classes in both sets of questions are shown in Table I.

TABLE I

AVERAGE SCORES AND AVERAGE DEVIATIONS THEREFROM, BY METHODS AND BY TRADE CLASSES

Trade Class	Method		No. of Cases
	I	II	
Novices	78.3 ±12.5	12.5 ±5.1	35
Apprentices	128.5 ±37.5	37.0 ±8.5	4
Journeymen	206.4 ±12.7	63.3 ±5.9	36
Experts	207.4 ±14.3	66.8 ±3.9	26
Possible Score	248	72	Total 101

SUMMARY

(1) A written trade examination, based on the multiple choice answer method, was given, without time limit, to 35 novices, 4 apprentices, and 62 journeymen.

(2) Two combinations of questions were tried—the complete set of 62 and a specially selected series of 18.

(3) Both combinations, with a very high degree of accuracy, separate the novices, apprentices and journeymen, but neither of them differentiate between the journeymen and the so-called experts, although the eighteen-question set is slightly more efficient.

(4) In the absence of extreme language difficulty, the written group examination can be used with great advantage to (1) eliminate the bluffer, (2) to separate the apprentice of average attainment from the journeymen.

The authors acknowledge the help accorded them by Messrs. T. J. Kirby, R. Pintner, and R. M. Reinhold, all members of the Army Trade Test Division.

MINOR STUDIES FROM THE PSYCHOLOGICAL LABORATORY OF INDIANA UNIVERSITY

III. The Relation of the General Intelligence of School Children to the Occupation of their Fathers

By S. L. Pressey and Ruth Ralston [1]

1. *Problem and Materials.* It has been generally recognized by workers with tests of intelligence—from Binet on—that results with such tests from school children may vary surprisingly from one neighborhood to another; children in one neighborhood may average as much as three years in "mental age" above children in another type of neighborhood.[2] It has usually been argued that such differences were but further evidence of the superior native ability of the people living in the more prosperous communities, and of the inherent incapacity of the poor. That is to say, a direct correlation has been thought to exist between the economic and sociological status of a neighborhood or locality and the intelligence of its people,— as measured in its school children.

Such community comparisons, by means of mental tests, are as a matter of fact, decidedly interesting, and have value both psychologically and sociologically. The test findings may, apparently, reflect local conditions with surprising detail.[3] It is obvious, however, that to select children by neighborhoods is to select them very roughly indeed; if we wish to study the relation between the mental endowment of school children and the economic and sociological status of their families we should select our cases according to some economic or social standard which can be applied to each particular child or family in question. Children might thus be grouped

[1] Help in the collecting of the data was also given by Mr. J. Ward Starr.

[2] See for instance W. Stern, *The Psychological Methods of Testing General Intelligence*, pp. 50-57; Terman, *The Measurement of Intelligence*, pp. 114-118; Yerkes, Bridges and Hardwick, *A Point Scale for Measuring General Intelligence*, pp. 73-83.

[3] See for instance, Pressey, S. L., A Comparison of Two Cities and their School Systems by Means of a Group Scale of Intelligence, *Journal of Educational Administration and Supervision*, Vol. V, 1919, pp. 53-62.

information, be of distinct value in the investigation of an individual's interests, and in sounding the extent to which he is acquainted with various types of experience — good or bad. (3) In the third place, it was hoped that the test might yield data of some value in connection with efforts to differentiate deteriorated individuals in work with the insane. Roughly it may be said that mental deterioration shows itself in a fund of knowledge much greater than would be accounted for by present capacity for acquirement of ideas. Such cases should, then, be distinguished by a relatively good record on tests of information along with very poor records on tests of learning. As an informational test to be used for this purpose such a list as has been just suggested should be more satisfactory for the general run of psychotic cases than Terman's more literary list. The intention was to investigate the value of the test for each one of these three purposes. It has not been possible for the writer to complete the study as he had planned because of the many pressing demands of work in other directions. But a description of the test and the tentative results thus far obtained may be of some interest.

The list of words is shown below. It was arrived at by a study as to which fields of information were most familiar to the general run of cases at the Psychopathic hospital, by in-

RANGE OF PRACTICAL INFORMATION

Moral	Religious	Governmental	Legal	Mechanical
Stealing	Creed	Mayor	Jury	Pulley
Lying	Disciple	Senate	Plaintiff	Plow
Gambling	Benediction	Constitution	Probation	Scythe
Cheating	Heretic	Tax	Larceny	Rudder
Pity	Sacrament	Tariff	Sheriff	Piston
Cruelty	Altar	Naturalization	Perjury	Forge
Revenge	Monastery	Consul	Embezzlement	Dynamo
Envy	Bishop	Regent	Blackmail	Insulation
Patriotism	Easter	Quorum	Forgery	Derrick
Duty	Trinity	Treaty	Bail	Semaphore

Industrial	Commercial	Recreational	Domestic	Public
Piece-work	Bank	Ace	Oven	Post-card
Floor-walker	Check	Dice	Yeast	Telegram
Corporation	Credit	Short-stop	Flue	Ambulance
Line-man	Pawn	Touch-down	Sirloin	Steerage
Broker	Mortgage	Jockey	Borax	Editor
Strike-breaker	Auction	Cueball	Thimble	Insurance
Boycott	Contract	Odds	Dimity	Vaudeville
Stoker	Bankruptcy	Lager	Mop	Taxi
Foreman	Discount	Stogy	Tape	Pension
Rake-off	Lease	Rag-time	Eaves	Militia

quiry of doctors, social workers — and also patients — there, and by newspaper reading. It is by no means as satisfactory as might be desired, and a double meaning for several of the words has been discovered; but it seemed unwise to try to improve it further without trial in definite form. The writer at first wavered between two methods in selecting the words. It was possible (1) to try and obtain words, within each field of information, which any intelligent person of fair general information should know. Or, (2) the list might be made up of words which only an individual especially conversant with that field would be able to define. If the list were made up in this last way, and used either as an information or a free association test, it might well have some detective value—might bring to light a hidden interest in unfortunate subjects, perhaps. With the formation of such a list in view the writer for some weeks carried a special pad of paper on which he jotted down all the highly special words which he could obtain from interviews with various unsavory characters appearing at the hospital, and with street-car men, clerks, sailors, mechanics, and others having their own special supply of odd words. But it was finally decided that the first concept was the better: the second basis of selection would lead to a list of very limited usefulness.

2. *Results.* As a test of general ability and a check on the results of Terman's vocabulary test (the use for which the list was first intended) the value of the list would best be demonstrated by data from children of the poorer classes or from adults of little schooling; such cases might be expected to do very poorly with the Terman words, but to give results with the writer's list much more in keeping with their mental age. The writer has been unable to obtain any systematic data of this sort, but has instead tried to investigate (a) the value of the list as it may be used with average school children, and (b) its usefulness as a means of studying differences in interest and special information. In investigating the first problem the Terman list of words, and the 100 words of the above list, were given as a group test to 56 children in the fifth grade in two schools in Bloomington, Indiana. It was desired to find out (1) the most satisfactory method of scoring, (2) the correlation of results from the two lists with some independent measure of general ability, and (3) the extent of the specific contribution made by each list to the correlation.

Two methods of scoring the results were used. In the first place, they were scored according to Terman's rule, one point being allowed for each word which was defined with a degree

of exactness sufficient to indicate some knowledge as to its meaning. That is, scoring was very lenient and no partial scores were allowed. The results were then gone over a second time, a score of 1 being given for a bare indication of acquaintance with the word, and two points being allowed when the word was correctly defined. The correlation of the results with the practical information test, scored in these two ways, was .86. The correlation for the Terman list was .84. These correlations, of course, suggest strongly that it makes little difference which method of scoring is used.

Results were now correlated with findings from a group scale of general intelligence given to these same children a few months previous.[3] The correlation of the practical information test, scored in the first way mentioned, with these ratings was .40. The correlation, scored in the second way, was .45. The results would suggest that the more exact method of scoring might be expected to give more significant results. The Terman list scored in this last way was now correlated with the scale. The correlation was .50. Apparently for school children (if we may generalize from these results) Terman's list is slightly better as a test of general intelligence; a result which surprised the writer, but of course, does not affect the question as to the relative merit of the two lists for the special purpose already mentioned.[4] A higher correlation in both instances would be expected from older children, since results from older children would be less affected by irrelevant cir-

[3] Pressey, S. L. and L. W. A Group Point Scale of General Intelligence. *Journal of Applied Psychology*, Vol. II, 1918, pp. 250-269.

[4] Whatever the comparative merit of the two tests for any of these purposes, the Practical Information test is at least vastly superior in the number of extraordinary ideas it brings to light. In the present experimentation the Terman list resulted in only one mild epigram: hysterics is "sick in your head." But with the writer's list such gems abound. To pawn is to "sell things temporarily;" credit is "to pay for something some other time;" a mortgage is "when another person can take what he wants off you." Larceny is "stealing things that don't amount to much;" cheating is "to take money from people who can't count good." The religious terms bring out vividly the prevailing cynicism among the younger generation: an altar is "an object on which sacrifices were made in old times—in use now;" and Easter "is the day on which Christ *was supposed to have risen!*" Two curious bits of etymology also appear here: a disciple is "when it is not coupled together," and a bishop is "the owner of a shop." Yeast is "put into bread to make it grow." A floor-walker "quiets the baby;" rag-time is "bouncy music." An editor "puts the news into the newspapers." A boycott is "a boy that has a uniform on and will be a soldier some day;" patriotism is "to be kind, and put out a garden." The finding of such choice bits gives a real zest to the labor of scoring.

cumstances such as differences in ease of writing. But since the children were allowed all the time they needed there is no reason to suppose that any special difficulty in writing or in expression played any very serious part in the results.

A mild effort at partial correlation resulted in a correlation of .78 between the two lists, r_{12} was found to be .26 and r_{13} was .11, the first variable being the group scale, the second being the practical information test, and the third the Terman vocabulary. Apparently the practical information test does make its own contribution to the correlation. Correlation with an individual examination such as the Binet, and oral definition for both lists, would, of course, be much more valuable. However, the above figures will give some rough idea as to what might be expected in general of the comparative correlation of each informational test with ratings of general ability.

An effort was next made to discover whether clear differences in knowledge of special fields might be found by means of the practical information test. Analysis by sex was first tried. In making this analysis data from the 56 fifth grade children already mentioned was combined with results from 29 eighth grade children. The median for the boys in each group of ten words, for each grade, was found, the number of girls above this median counted, and the total number of girls above the median expressed as a per cent of all the girls tested. The results were as follows:

Word Group—	1	2	3	4	5	6	7	8	9	10
Per cent girls above boys median	50	50	25	17	20	25	50	20	61	58

There are here clean-cut sex differences. The girls are distinctly superior in knowledge of the "domestic group" of words, and slightly superior in the list having to do with public utilities and information of that same general nature. Boys are superior in knowledge of the "governmental" words, of the legal and mechanical words, and also of those having to do with recreation.

A comparison of the fifth and eighth grade was also tried. But it was impossible to find any definite difference in increase of knowledge along different lines. The children in a small colored school were next given the list, a total of 36 children in all. Of these 14 were in the fifth grade, 15 in the sixth, 6 in the seventh and 1 in the eighth; the group was, therefore, lumped and compared with the results from the fifth grades. The per cents of colored children scoring above the median

for the white children, in the ten groups of words, were as follows:

	1	2	3	4	5	6	7	8	9	10
Per Cent:	45	56	81	56	19	80	81	81	72	80

To the writer, at least, the results are distinctly surprising. The colored children average about 1 grade above the white chidren and over two years older. Nevertheless, they show a poorer average knowledge of moral terms, about the same acquaintance with religious terms. They show what might almost be called a total lack of acquaintance with the general vocabulary of mechanics. With the small number of cases these results should, of course, not be made too much of; it should be remembered also that the words are graded in only the most rough fashion and the test otherwise loosely organized. But it may at least be said that the results are suggestive. The writer cannot but believe that the races differ in the tendency of their interests; and he feels decidedly that a valuable way to investigate such a matter is to measure the information acquired along various lines.[5]

3. *Discussion.* A word remains to be said on these few findings. They surely suggest fundamental differences, as regards information and interest, between the sexes and between the two races. The writer feels strongly that until we know more about differences in interests we cannot know what we should with regard to sex and race differences. The tendency is to investigate such matters by somewhat isoteric and roundabout methods. We use the free association test, for example, act mysterious, employ a stop-watch, and otherwise embarrass the subject. A straight-forward investigation of information might very well, in such instances, bring the same data with no difficulties of cooperation. As one method of investigation the writer would like to suggest such lists as those above. In its present form the test here described is

[5] It should be stated that in making these analyses the second method of scoring was used, that is, a child might get a zero, one, or two, on each word. As indicative of the general fall of the figures it may be added that the median for the fifth grade, when the practical information test was thus scored, was 42.5; for the eighth grade it was 106 (out of a total possible of 200 points). The fifth grade medians for the separate groups were as follows:

	1	2	3	4	5	6	7	8	9	10
	12	3	2	1	5	0	4	0	8	5
For the Eighth:	18	11	12	5	11	6	12	9	14	17

The fifth grade was composed of 37 girls and 29 boys; the eighth grade group of 14 girls and 16 boys.

obviously not satisfactory; it is not adequately organized, it is hard to score and awkward to handle; but it seems to have very good possibilities as a method of investigation. It may have a practical value too. But the writer has been diverted from that matter by the research interest of the data. He hopes to have further material to offer later.

4. *Summary.* The paper may be briefly summarized;

1. It is argued that an "unselected" vocabulary test is unfair to delinquents and to adult cases of little education. As a substitute a selected vocabulary covering certain fields of common information is offered.

2. A comparison of this test with Terman's vocabulary test is made. It is found that fifth grade children give results, correlating more highly with findings from a group scale of intelligence, with the Terman vocabulary test than with the writer's information test.

3. Distinct sex and race differences in the various fields of information are found.

4. It is suggested that such a test offers a convenient method for the investigation of sex and racial differences in interest.

V. The "Efficiency" of a Group Scale of Intelligence in Prognosticating Success and Failure in Junior High School

By S. L. Pressey

1. *Problem and Materials.* The paper deals with the prognostic value of a group scale of intelligence previously described in this *Journal*,[1] for forecasting success or failure in Junior High School. This scale was given in March, 1918, to all the children in the sixth grade (the entering grade) of a large Junior High school—a total of 165 cases. In the fall of 1919 the subsequent school history of these cases was investigated. Of the original number 20 cases were found to be lost to the study, because the family of the child had moved away, or for other similar reasons having no connection with school standing (cases leaving school because of school failure have been included in the study, as will be described below). For each one of the remaining 145 children the school marks for June, 1918, January (semesters) 1919, and June 1919 were obtained. The paper has to do with the relation between these marks and the test scores.

Two questions have been put: (1) What test scores were made by those children who failed in school (that is, using the school's own definition, those children who obtained marks below 75 in three or more of the four "promotion subjects"—arithmetic, geography, history, language)? (2) What is the relation of test scores to later success in school as shown by the last available marks (June, 1919), or when this class as a whole had completed the seventh grade? In each instance the effort has been to present the findings in terms that would be readily interpretable for use in individual diagnosis; in fact the interest of the writer has been rather in these methods than in the particular results dealt with.

2. *Results:* (*a*) *Prognosis of School Failure.* For the purpose of studying the relation between test score and failure in school the distribution of test scores for the 145 cases was divided into ten parts by calculating the deciles throughout.

[1] Pressey, S. L. and L. W. A Group Point Scale for Measuring General Intelligence, *Journal of Applied Psychology*, Vol. II, 1918, pp. 269-287.

The number of failures scoring in each decile was then determined. The results run as below. As has already been said, the tests were given in March, 1918. The first row below gives the number and distribution of failures for June, 1918; the second row gives the same facts for January, 1919, and the third row the data for June, 1919. The last row summarizes by giving the position of the scores of all the children who failed during any one of these three terms.

	Decile Division										Total
	1	2	3	4	5	6	7	8	9	10	
Failed:											
June 1918	9	6	3	1	2	0	1	1	0	0	23
Jan. 1919	6	4	3	1	1	0	0	1	2	0	18
June 1919	9	6	2	4	1	0	0	2	1	1	26
All failures	11	9	5	4	2	0	1	2	2	1	37

It should be added that this final row does not count any child twice, but shows the number failing at any one or all of these three terms. It should also be said that the 1919 rows include as failures those who failed in June, 1918 and left school because of the failure.

Perhaps the most striking thing about these results is the bimodal distribution of the failures. The vast majority of the children who fail tested well at the bottom of the distribution. But there is another small group near the top. There is a suggestion here of a maladjustment not only between dull child and school, but between school and superior child as well — perhaps some of the brighter children do not find sufficient stimulus to their abilities in the routine school work of their "normal grade." Anyhow, this little group of failures who test as above average in ability is distinctly interesting.

The matter is, however, somewhat aside from the problem in hand. The practical problem is this: What indication can be obtained, if such an examination be given to children in their first year in Junior High School, as to those pupils who will later fail in their school work? For such a practical diagnostic purpose the writer would like to suggest the following table, which is simply the last row of the table above, expressed in terms of per cent of the number of cases in each percentile division who were failures.

	Decile Division									
	1	2	3	4	5	6	7	8	9	10
Per cent. failing, or—per cent probability that a given case will fail	76	62	35	28	14	0	7	14	14	7

Thus there were 11 failures in the first decile division; there were a total of 14.5 cases in each decile division,—and 11 is 76% of 14.5. The other per cents were similarly found. The table shows the incidence of failure throughout the entire distribution of test scores. It is also, however, readily interpretable for purposes of individual diagnosis. Suppose a principal gives the scale to the entering class in his Junior High School during their first year. He may assume (if the school from which the writer's data is taken may be considered reasonably representative) that if a given child scores below the ten percentile for the grade there is a 76% probability that that child will, within the next two years, fail at least in one semester of work. If another child scores between the thirtieth and fortieth percentile there is a 28% probability that that child will fail — and so on for the other decile divisions. The method is direct and obvious, and would seem of some general usefulness.[2]

(*b*) *Prognosis of School Attainment as Measured by School Marks.* In order to obtain some more general indication as to the relation of test scores to future school accomplishment the relation of these scores to the marks obtained by all the children in June, 1919 — almost a year and a half later —was studied. The median mark of the marks obtained, at this time, by each child, in the five subjects arithmetic, reading, grammar, history, and geography, was used to represent school success. And a correlation was run between these median marks and test scores. The correlation was found to be .50.

However,—as has been well pointed out recently[3]—a correlation coefficient is a very inadequate way of presenting such data for purposes of individual diagnosis. In order to obtain a statement more satisfactory, for such a purpose, the scatter diagram from which the correlation was plotted was divided into five parts each way, by ruling lines through the 20, 40, 60, and 80 percentiles of each distribution. And the table was made to read as for 100 cases, by expressing the number in each compartment of the diagram as a per cent of 145—the total number of cases. The result was as follows: (I being best and V poorest in each array):

[2] A somewhat different application, and to a psychiatric problem, has been made in a previous article, "Irregularity as a Measure of Deterioration" *Jour. of Abnormal Psychology*, December, 1918.

[3] Thurstone, L. L. "Mental Tests for College Entrance," *Jour. of Educ. Psychol.*, Vol. X, 1919, pp. 129-42.

DIVISION INTO FIFTHS—TEST SCORES

		V	IV	III	II	I
Division into fifths—school marks	I	0	0	2	5	13
	II	1	4	4	8	3
	III	4	4	5	4	3
	IV	7	6	5	1	1
	V	8	6	4	2	0

If, now, the hypothetical principal above mentioned has given the scale to his sixth grade, and finds that a particular "problem case" scores in the upper fifth of the class, he may conclude that the probability is 13 out of 20 that that child will, at the end of the next school year, be receiving marks which place him in the upper fifth of his class. There is only 1 chance in 20 that this child will be marked below the middle group. There are 0 chances out of 20 that the pupil will appear in the poorest fifth, in school grades,—and so on.[4] The method is direct, readily understood by teacher or principal; and not only the probable diagnosis, but also something as to the definiteness of the indication, is shown.

As with the previous method, a larger number of cases is necessary to make the table reasonably reliable—the ideal form would probably involve decile divisions, the table then reading as for 1000 cases, or 100 in each array. But the writer feels that the methods themselves should be of distinct practical value, particularly in dealing with the very concrete and definite prognostications of business and vocational psychology.

3. *Discussion.* A word remains to be said with regard to such use of school marks, and school failure, as criteria in terms of which to evaluate tests of "general intelligence." It might be argued that the peculiar value of such tests was in in their ability to get beyond the particular narrow set of facts by which school marks are usually determined. And the ultimate criterion becomes—what the child *would* do, in a system highly adaptable to individual differences. But we also may wish to know what children, *things being as they are,* will fail next year, a year later; we may wish to know the children now in the eighth grade who will fail in high school

[4] It should be said that eight cases leaving school because of failure in June, 1918 have not been included in the correlation. Strictly speaking, they should be, somehow, probably at the bottom of the distribution of marks; our hypothetical principal does not know in advance who is going to leave school and should be provided with a prognosis table, covering *all* the cases under consideration But for present purposes the omission of these eight cases may be considered negligible.

should they go there, the college entrants who will fail their Freshman work. It may very well be that for prognosis over a brief period, in a situation such as is dealt with in the present paper, a set of achievement tests would give a more definite indication than tests of more general content; the writer hopes soon to present data bearing on this point. The study of such very concrete and rigidly defined problems is certainly much needed; such study would appear the only empirical and scientific way of determining the validity of the concept of "general intelligence"—that is, the validity of the theory that prognosis problems can be generalized.

4. *Summary.* 1. The paper deals with scores obtained from the sixth grade children of a large Junior High school, and with the school history of these children for three semesters since the tests were given. A definite statement of the prognostic value of the tests, over this period, is sought.

2. A numerical statement is obtained of the probability of failure in school implied by scores at various points on the distribution of test scores.

3. An analogous statement for prognosis of marks, over one year, is presented.

BOOK REVIEWS

JAMES BURT MINER. *Deficiency and Delinquency, An Interpretation of Mental Testing.* Warwick and York, Baltimore, 1918, p. 352.

The author here presents a comprehensive study of mental testing as applied to deficients and delinquents, assembling from a vast body of literature the various conclusions thereto appertaining. As the title suggests, the study is an interpretative one. We may add that it is also a wholesome critique. His claim to greater practicality in this important field of psychology is verified throughout, as there is shown to be a place for the serious attempts of social welfare workers, teachers, and amateurs, as well as the trained clinician.

Heretofore, the great complexity of the problems of deficiency has not been adequately solved by the employment of objective or qualitative tests alone, since the results have always been vitiated by certain errors necessarily connected with examinations. While the earlier objective tests have contributed much to basic data, and while they are sufficiently accurate to serve as a check to empirical judgment respecting obvious cases of deficiency, they are relatively worthless for the diagnosis of that larger group of borderlines, the accurate testing of which alone determines the true value of a test. The central theme and purpose of the study is, therefore, to set forth and explain the requisites of a system of tests which will serve to diagnose the troublesome "borderline" cases. The need of redefining the borderlines of deficiency becomes obvious to the author through practical experience with deficients in the Minneapolis Juvenile Court. Valuable data is also gathered from the Glen Lake Farm School for juvenile delinquents. As a further background for this study, the author has had the advantage of visiting scores of institutions in Austria, England, Switzerland, Germany, France, and Italy. To all this may be added the author's knowledge gained through the study of law, and the legal problems now fast becoming coextensive with social problems.

Accepting the common conception of feeble-mindedness, which holds it to be a lack of development rendering the individual socially deficient and requiring constant supervision or guidance, and excluding from this category the so-called "conative" cases of deficiency, the author's contribution may be summarized as follows:

1. Deficiency and delinquency are not mutually inclusive. Deficiency may be regarded as a state of permanent undevelopment, while delinquency may or may not accompany it. In order to eradicate subjective opinion in the differentiation of the two, a conservative borderline on the basis of Binet or other tests, should first be established, so that treatment of the delinquent is assured a still further individual consideration.

2. Present-day tests are inadequate for reliable diagnostic purposes, since they tend to exaggerate qualitative differences among abnormal subjects. Deficiency may now properly be looked upon as a relative and "quantitative" condition; every degree and shade of undevelopment is the rule.

3. The number of defectives in proportion to the school and general populations have been variously estimated; and these estimates indicate

something of the illusory and unscientific nature of tests and their *interpretation*. Therefore a "*percentage definition* of tested deficiency is suggested for determining the borderline below which an individual may be presumed to be so deficient as to justify isolation." The most conservative estimate of the number of defectives may now be said to be: "Those who would at 15 years of age be in the lowest 0.5% in tested ability among a randomly selected group, may be presumed to be so deficient as to justify isolation. Above them the next 1.0% may be regarded as uncertain, since the bulk of them would require some *supervision or guardianship during life*." The *verification* of such borderlines may be regarded as the main contribution yielded from the new data here studied.

4. Practically speaking, the location of borderlines on a test scale is symptomatic only, since different communities offer different results.

5. Emphasis should be placed on the social significance of deficiency both in its diagnosis and ultimate results. Deficiency and delinquency should therefore be referred to a commission consisting of expert *medical and psychological assistants, so that conative cases may not* be unjustly retained in public institutions permanently. To this end, a "temporary home school" is advisable.

6. Legal definitions of the criminality of deficients should be so changed as to obviate all subtle questions of mental analysis, such as: "was laboring under such a defect of reason as not to know the nature and quality of the act he was doing or know the nature of the wrong" (New York Law). Such law places too much emphasis upon the *intellectuality of the accused, whereas in most cases of delinquency, the* crime is more closely related to the emotions. The problem of criminality is a "problem in the treatment of hereditary diathesis in which the mental deficiency is the largest factor."

7. Whether society should adopt a policy of sterilization, or that of isolation, by force or consent, is an open question. At present it appears to be a question which may well await the perfecting of developmental tests that may assure accurate diagnosis. But wherever the cases are obviously permanently deficient, isolation is advisable for girls and women of child-bearing age.

8. External influences, such as criminality in the home, character and amount of education, *standards of living*, etc., *seem to be less important* than is heredity.

9. The schools have in all cases of deficiency, a special work to do in the discovery of such pupils, and in the referring of the same to the proper authorities. The school may also do a great deal in gathering data relative to the various correlations of conduct and mental attainment, which so far have been relatively unscientific.

University of Vermont. J. W. Sprowls.

Sherwin Cody. *Commercial Tests and How to Use Them*. School Efficiency Monographs. World Book Co., 1919.

The author presents valuable tests for determining ability to perform fundamental operations in business offices, viz., tabulation, reproduction and following of instructions, elementary *invoicing*, elementary arithmetic, elementary English, business letter writing, stenography, typewriting, penmanship and filing. These tests are in line with the tendency to test specific abilities and have the further qualification of being short. They have been standardized by use among thousands of grammar school and high school graduates and office employees of several large business houses, the average performances in the several

tests being given for each group. However, in giving these norms the author fails to add any measure of variability, thus detracting materially from the scientific and practical value of his work.

The author's use of the expression "business ability" to describe these tests is possibly misleading. They do not test business *capacity* but rather achievement and are therefore classifiable with educational or proficiency tests. Used alone they have rather limited value for the prospective employer, since they give no indication of an individual's future value as an employee nor of his ultimate vocational fitness for a business position. General intelligence, learning ability, habits of industry, etc., must be obtained from other sources or be assumed to correlate highly with the scores in the tests. However, improvement rate in a series of tests, together with school marks and teachers' ratings, could be used successfully by employers for selecting employees who have a high degree of capacity.

One test in the series, the Tabulation Test, is called an 'intelligence' or 'mental alertness' test. This amounts to stretching the conception of intelligence tests rather severely, but the statement is in line with the author's contention that a "high degree of accuracy in figures, ..., is an evidence of mental capacity," and shows "a discipline in mental control which must carry over into all the activities of life." Even if this statement is accepted it cannot be assumed that the reverse is equally true, for high mental capacity and intelligence is possible and frequent where ability to use figures with speed and accuracy is lacking.

An outstanding feature of the book is the method used for grading and checking the tests in the schools. The pupils themselves do the checking. This makes for speed and, in the author's opinion, gives a sufficiently high degree of accuracy. For example in New York, in 1917, "the tests were given to 6,000 eight-grade pupils, three tests to each. The margin of points overlooked or wrongly marked was not sufficient to change the median for any class as much as one per cent."

A gross error appears in the form of directions for giving the tests (page 57): "Tell pupils they must be as careful not to get too high a mark as not too get too low a mark, for then on the next test they will not show the improvement which is equally important with a high grade to start with."[1]

Carnegie Institute of Technology, Pittsburgh, Pa. H. C. KENAGY.

WILLIAM FRETZ KEMBLE. *Choosing Employees by Mental and Physical Tests.* The Engineering Magazine Company, New York, 1917, p. xiii+333.

The purpose and claim of the book is perhaps best expressed in the author's own words: "The following pages reproduce a series of tests which not only will pick out the man of great capacity from among a mixed company undergoing the test, but will arrange a staff of men in the approximate position of their ability when certain weights are assigned according to the nature of their work." This claim seems totally unsupported by results. Kemble not only does not prove in any way that his tests will do what he claims for them if handled by an experienced man; but he gives no accurate and sufficiently detailed instructions; therefore if the tests were given by an inexperienced man, there would be no hope of their receiving a proper evaluation even were the tests adequate, which we believe they are not.

[1] Review approved by the Committee of the American Psychological Association on Publications in Applied Psychology.

For example, Kemble nowhere states how he has standardized his tests. In several cases he gives the "human average" but says nothing of how this average was obtained. He suggests a psychogram of the subject, but does not indicate how this psychogram can be evaluated by the business man. And what is equally important, he does not seem to recognize any limitations to tests, nor does he differentiate between the precision of different tests, being willing to give an exact *score in spelling ability and an equally exact one in honesty.*

In order to so estimate a man's ability, Kemble would use four sources of information. The inadequacy of these is apparent to any one who understands the laws of testing. They follow:

I. An employment blank, consisting of five parts:
   a) a personal history blank
   b) a financial statement
   c) an expenditure blank
   d) a questionnaire concerning the religious and moral life of the person
   e) a personal belief blank

II. The "Peg Board Tests."

III. A series of mental activity tests. There are twenty-one of these tests, most of them of the ordinary type: memory for sense material, etc. However, a few are of rather an exceptional nature. The general information test consists of a series of questions, the information for answering which would be most apt to be picked up among day laborers—for example: "Give the proportions used in making mortar." The ethical test has questions similar to the following: "If you were about to take a trolley and a friend got off a car and offered you a transfer, would you use it?" The personal opinion test is rated in an original way; the answers to the questions are decided by the vote of at least ten high-class executives. The test then is supposed to differentiate between the executive and the popular type of mind. One's honesty is rated by the number of the following questions that he answers by "Yes": "Do you smoke? Do you drink? Do you use slang? Do you swear?" Kemble's method of using these tests is to decide what elementary qualities are involved in any occupation for which he is testing applicants, pick the tests which test these qualities and sum them for the applicant's score for that occupation. Anyone at all familiar with testing realizes at once the danger of such a proceeding.

IV. A study of the physical characteristics of the subject, type of face, length of fingers, etc.

In general, Kemble has made a rather unique collection of tests. A few have new features which perhaps offer ideas, but his claim of 90 per cent accuracy and the lack of standardization makes the book of little value for practical purposes, and even pernicious if used by an inexperienced person.[1]

Carnegie Institute of Technology, Pittsburgh, Pa. M. A. BILLS.

FREDERICK SCHLEITER. *Religion and Culture, A Critical Survey of Methods of Approach to Religious Phenomena.* Columbia University Press, New York, 1919, p. 206.

The subtitle of the book should be borne in mind when one is reading it. It is the methods of approach to religious phenomena that the author is discussing. And the first method taken up is that propounded

[1] Review approved by the Committee of the American Psychological Association on Publications in Applied Psychology.

by Buckle, the studying of limited areas intensively and then drawing general conclusions from the results obtained. This method "bristles with fallacies and insupportable presuppositions, the most obvious and far reaching of which embodies the idea that the ethnographic phenomena found in some specially selected area, are the result of an independent development, and constitute, as it were, an indigenous entity possessing complete historic individuality" (p. 10).

The comparative method which consists "in bunching of static facts to indicate the common properties of primitive mind," . . . "has served to bring together similar customs, rites and ideas the world over, to stimulate investigation of them and has, in many cases, laid the basis for their elucidation" (pp. 21, 27). But it should be remembered that all genealogical hypotheses "lack objective confirmation and that their relative merit is to be judged largely on the ground of obvious plausibility and ingenuity" (p. 40).

The framing of ethnographic analoga the author characterizes as "a somewhat romantic procedure" (p. 67).

Animism and the mana theorists are equally incompetent to solve the problem, since the former "is far from literally true" (p. 80), and the latter "are full of inconsistencies, obscurities, and ambiguities" (p. 123). The discussions of animism and mana are very interesting and instructive.

The rôle that the concept of causality has played in religion and culture is truly great. This very obscure and ambiguous concept has lately come in for a great deal of criticism, and some scholars have gone so far as to suggest that it be rejected altogether. Mach says: "There is no cause nor effect in nature; nature simply has but an individual existence; nature simply is" (p. 142). Russell, who is perhaps one of the most severe critics, points out that "in advanced sciences such as gravitational astronomy, the word 'cause' never occurs" and that the "concept is not employed at all nor does it find embodiment or asylum in the algebraic equation" (p. 144). "Fiske characterized the development of science as a progressive 'de-anthropomorphization' and Comte supposed that the evolution of society involves the continuous elimination of 'fetishistic' elements" (p. 146).

Lastly the author discusses "the convergence of mental processes which lead to the attribution or predication of a dynamic relation between two or more elements" (p. 153). In this connection he calls attention to Dresslar's work on superstition among persons "professionally trained for the work of teaching." This work shows that 44.9 per cent believed either wholly or partially in certain superstitions. But "neither the character of the connecting link nor the manner of the achievement of the results appears in the tables" (p. 157). In another connection the author points out that "in many cases, he does not know precisely how these events, objects or processes are connected,—in other words, the nexus does not rise into his consciousness" (p. 161).

But there is not only an unconscious acceptance of the articulating mechanism, there is also to be found "a highly conscious and rationalistic enquiry in which the mind attempts to grasp, by acts of deliberate apprehensions, the causes of an event and then launches out boldly upon a more or less boundless path" (p. 164). Several interesting illustrations are here produced, both among primitive men and men of our own time, all of which go to show that an "enormous range and diversity of mental processes are involved in causation" (p. 191).

The purpose of the book is to "suggest that these logical processes,

which lift facts out of their indigenous habitats and deal with them in their universal aspects, should be applied with more critical caution than is customary" (p. 193). The positive contribution of the book is naturally not as great as the negative. But it constitutes a good critical survey of the field it covers and is therefore a valuable contribution to the literature on this subject.

KARL J. KARLSON.

The following publications have been received:[1]

JAMES WINFRED BRIDGES. *An Outline of Abnormal Psychology.* R. G. Adams and Company, Columbus, O., 1919.

W. J. CRAWFORD. *Experiments in Psychical Science, Levitation, Contact, and the Direct Voice.* E. P. Dutton and Company, New York, 1919.

L. M. GILBRETH. *The Psychology of Management.* Sturgis & Walton Company, New York, 1918. Reprinted without changes from 1914.

HENRY C. LINK. *Employment Psychology.* The MacMillan Company, New York, 1919.

NORAH H. MARCH. *Towards Racial Health.* E. P. Dutton & Co., New York, 1919.

DANIEL STARCH. *Educational Psychology.* The MacMillan Company, New York, 1919.

LEWIS M. TERMAN. *The Intelligence of School Children.* Houghton Mifflin Company, Boston, 1919.

MARGARET J. HAMILTON. *Psychological Analysis and Re-education, with Case Studies.* Reprinted from *The Journal of Educational Psychology*, Feb., 1919.

M. LAIGNAL-LAVASTINE. *The Internal Secretions and the Nervous System.* Trsl. by F. T. Robeson. Reprinted from *Nervous and Mental Disease Monograph Series.* No. 30, 1919.

S. D. PORTEUS. *Cephalometry of Feeble-Minded.* Reprinted from the Vineland, N. J., *Training School Bulletin*, June, 1919.

——— ———. *Porteus Tests—The Vineland Revision.* Reprinted from the Vineland, N. J., *Training School Bulletin*, No. 16, September, 1919.

ARTHUR H. SUTHERLAND. *Ungraded Rooms in Los Angeles City Schools.* First Annual Report, Division of Psychology, Year ending June 30, 1919.

J. E. WALLACE WALLIN. *Meeting the Needs of the Mentally Handicapped Child in School.* Reprinted from *Ohio Bulletin of Charities and Correction*, June, 1919.

---

[1] Mention here does not preclude further comment

——— ———. *The Field of the Clinical Psychologist and the Kind of Training Needed by the Psychological Examiner.* Reprinted from *School and Society*, April 19, 1919.

——— ———. *The Achievement of Mental Defectives in Standardized Educational Tests.* Reprinted from *School and Society*, Aug. 29, 1919.

J. HAROLD WILLIAMS. *The Intelligence of the Delinquent Boy.* Reprinted from *Journal of Delinquency*, Monograph No. 1, 1919.

CARRY ON, A Magazine on the Reconstruction of Disabled Soldiers and Sailors, Vol. 1, No. 10, July, 1919 (probably the last isssue).

REVISTA DE PSIQUIATRIA. Lima, Peru, Vol. 2, No. 1, July, 1919.

REVISTA DE EDUCACION. Santo Domingo, R. D. Vol. 1, Nos. 3 and 4, 1919.

NOTES

A special committee of the National Research Council, consisting of Dr. R. M. Yerkes, chairman, and Dr. M. E. Haggerty of the University of Minnesota, Dr. L. M. Terman of Stanford University, Dr. E. L. Thorndike of Teachers College, Columbia University, and Dr. G. M. Whipple of the University of Michigan, with financial support from the General Education Board, have formulated a plan for using the army mental tests in schools. Such intelligence tests have been used in schools for some time on individual children, but the new plan provides for handling them in groups, even whole class-rooms at a time. The committee selected about twenty tests for careful trial. This trial was made on five thousand children. As a result the committee has now been able to select from the tests two series which seem to be the most satisfactory and these will now be tried on several thousand more children in order that they may be further perfected before they are finally offered to the teachers of the country for general use.

This carefully worked out program for group tests will make it possible and practicable to make wholesale surveys of schools annually, or even semi-annually, so that grade classification and individual educational treatment can be adjusted with desirable frequency. It is expected that the methods will be ready to be published for general use early in 1920. The army tests on which these new group tests for children are based and which were used with striking success and advantage during the war, were originally devised by a group of psychologists working under the auspices of the National Research Council.

The following statements are extracted from a report by Vernon Kellogg of Leland Stanford Jr. University published in *The Nation's Business*, November, 1919.

"The lesson of the war as regards science, then, is first, that the efforts of scientific men can be greatly speeded up by a proper stimulus; second, that these efforts can be made immensely more effective and fruitful by a proper organization; and, third, that while such an organization can be encouraged and helped by the Government, it can be effected by cooperative effort among scientific men themselves.

"The founding of the National Research Council is the outgrowth of this lesson. And the special characteristic—and a truly American one—which distinguishes it from the other rather similar organizations of England, Canada, Australia and Japan, and also from the already long-established government scientific bureaus of our own country is that, although officially recognized by the Government, it was not initiated or organized by it and is not supported by it. It is the outcome of a nation-wide cooperative effort of the scientific men of America, including representatives of the universities, the various special privately endowed research institutions, the scientific laboratories of the great industries and the Government scientific bureaus, and the numerous unattached specialists. And it is entirely controlled, in a thoroughly democratic way, by these many cooperating scientific

investigators. It derives its support from funds contributed from private sources.

The Council is organized primarily as a permanently constituted part of the National Academy of Science.

"Effective prosecution of the Council's work requires not only the cordial collaboration of scientific and technical men of the universities and special research institutions all over the country, but also those connected with the scientific and technical branches of the Government.

"As now organized, the National Research Council has permanent headquarters in Washington, with an executive staff of scientific men giving their whole time to the work of their respective positions. The Council's field of activities is divided among thirteen divisions, gathered into two main groups; first, one of six "general relations" divisions, and second, one of seven divisions devoted to special lines of science and technology. Each of these divisions has a resident chairman and a small office staff in Washington and a number of non-resident members.

"There is a Division of Anthropology and Psychology, as yet without a chairman,* but with a list of members chosen from among the leading psychologists and anthropologists of the country."

Dr. W. V. Bingham, Head of the Division of Applied Psychology of the Carnegie Institute of Technology at Pittsburgh, has been appointed Chairman of the Division of Anthropology and Psychology of the National Research Council. Dr. Bingham is an authority on methods for measuring the intelligence of normal adults. Early in the war, as Secretary of the Committee on Classification of Personnel of the Army, Dr. Bingham applied these methods to testing the mental capacity and fitness of recruits as a basis for assignment and training for particular military duties, and later continued this work as Lieutenant-Colonel in the Personnel Branch of the General Staff.

At present Dr. Bingham is connected with several bureaus of the Carnegie Institute, which are engaged in studying the application of these principles in commercial and industrial occupations. One of these bureaus, that of Personnel Research, is supported by annual contributions from 30 corporations. This bureau is engaged in ascertaining the best methods for selecting and developing executives, salesmen, and clerks.

Another bureau is applying the same principles in developing methods for selecting and thoroughly training workers in about 900 positions in seven of the leading department stores of Pittsburgh. These stores contribute $32,000 annually for these investigations. The financial support given this work of investigating and applying scientific mental tests reveals the confidence which is being placed by corporations and store executives in the money value of a rational study of their employment problems.

An appropriation of fifty thousand dollars in five annual payments has been granted to the Iowa Child Welfare Research Station of the State University of Iowa, of which Dr. Bird T. Baldwin is director, by the National Women's Christian Temperance Union, for the purpose of making detailed scientific investigation into the development and conservation of child life. The gift is for research and the station will

* Since this statement was first published, such appointment has been made. See notice following.

be given a wide range of freedom in the selection of its problems, which will center around the field of eugenics, nutrition, and the mental and social development of the child. Dean Carl E. Seashore is chairman of the advisory committee.

The Society for American Fellowships in French Universities offers each year for open competition among graduates of American colleges and other suitably qualified candidates a number of fellowships, not to exceed twenty-five, for the purpose of encouraging advanced study and research in French Universities.

The fellowships, of the annual value of $1000, are granted for one year and are renewable for a second year. They may be awarded in the following fields of study:

Anthropology, Archaeology and History of Art, Astronomy, Botany and Agriculture, Chemistry, Criminology, Education, Engineering, Geography, Geology, History, Law, Mathematics, Medicine, Philology: *Classical, Romance, Oriental, Semitic, English,*—Philosophy, Physics,—Political Science: Economics and International Law,—Psychology, Religion, Sociology, Zoology.

Fellows will be required to sail to France not later than July 1st of the year in which the award is made, to matriculate in a French University for the following session, and to pursue studies in the field of science designated in their awards. They will be expected to send accounts of their studies together with reports of their progress from their instructors.

Applicants, men or women, must be citizens of the United States and at least twenty years of age. They must be:

1. Graduates of a college requiring four years of study for a degree, based on fourteen units of high school work; or,

2. Graduates of a professional school requiring three years of study for a degree; or,

3. If not qualified in either of these ways, must be twenty-four years of age and have spent five years in an industrial establishment in work requiring technical skill.

Applicants must be of good moral character and intellectual ability, and must have a practical ability to use French books.

Applications must be made on Application Blanks furnished by the Society and must be accompanied by:

1. A *Certificate of birth*;
2. A Certificate of Naturalization, if needed;
3. A Certificate of College Studies;
4. A Certificate of Industrial Work, if needed;
5. Testimonials to Moral Character and Intellectual Ability;
6. A Photograph, signed and taken within a year; and
7. Printed or written articles, theses and books, written or published by the applicant.

Applications should reach the Secretary of the Society not later than January 1st of the year in which the award is to be made.

Application Blanks and further information about the fellowships may be obtained from the Secretary,

DR. I. L. KANDEL,
576 Fifth Avenue, New York.

The Massachusetts Society for Mental Hygiene, of which Professor William H. Burnham of Clark University is the president, held a Conference on Mental Hygiene in Worcester, Mass., November 14th, 1919. Dr. Walter S. Fernald, Superintendent of the Massachusetts School for the Feeble-minded, outlined a state program for the care of these unfortunates which involves a statistical survey and official record of all feeble-minded in the state, supervision and care of those who are not inmates of institutions, with the help of social and charity organizations, provision for special training during the formative period, and the inauguration of local mental clinics. Dr. Burnham showed that success and failure in school were important conditions influencing mental health and pointed out that present day administration and methods presented too many opportunities for failure and did not sufficiently utilize the mental stimulus of success. Dr. E. E. Southard, Director of the Massachusetts State Psychiatric Institute in Boston, summarized and classified the most important causes of mental ills under the headings of disease, ignorance, vice, legal entanglements, and poverty, which most frequently operate in pairs but only rarely alone, and showed that successful attack upon one member of a pair usually results in a great diminution, if not entire disappearance, of the other member. Dr. A. W. Stearns concluded the program by a brief summary of the aims and methods of the Society.

It is a matter of common knowledge that the increased cost of production of to-day is in large measure due to the demand on the part of labor for shorter working hours and more pay. The common way of the manufacturer of meeting this increased cost has been to pass it all and a little more on to the consumer by greatly raising the price of his commodity.

At the Athenaeum Press, where Ginn and Company's books are manufactured, there has just been inaugurated a plan of training foremen, forewomen, and other factory supervisors in production methods. A three-months' course is being conducted by the Business Training Corporation of New York for a group of forty-six, composed of the treasurer, press manager, department heads, and all foremen from the man in charge of a room of fifty to the one who supervises the work of two or three. The work consists in the study of especially prepared subjects, the solution of problems relating to each subject, and the discussion of this material at six bi-weekly meetings. At each meeting a lecture is delivered by an experienced production man. Afterward the meeting is thrown open to discussion of the material set forth in the text, and problems which may be peculiar to any particular branch of work in the plant. The subjects covered are teamwork, handling men, machinery and materials, organization, production records and management.

Beside a more thorough all-round knowledge of production methods, the course gives an opportunity for the leaders in the Press to get together, discuss important problems, not only among themselves, but with the higher executives, and to acquire at first hand the spirit of unity and team play which in the last analysis means the elimination of unrest.

INDEX OF SUBJECTS

(The subject headings of all contributions except Book Reviews and Notes are printed in SMALL CAPITALS.)

ABSOLUTE INTELLIGENCE SCALE 118
Accuracy, Test for Mental 179
ADULTS, AVERAGE MENTAL AGE OF 317
AIR SERVICE TESTS 103
AMERICANIZATION IN INDUSTRIES, CONFERENCE ON 269
ANALYSIS OF TRADE ABILITY 352
Applied Eugenics 96
ARMY PERSONNEL WORK 1
ATTITUDES IN DEBATE, MENTAL 156
AVERAGE MENTAL AGE OF ADULTS 317

Child, Kingdom of the 100
CHILDREN IN GOOD AND POOR FARMING DISTRICTS COMPARED, 283; COMPARISON OF COLORED AND WHITE, 277; THEIR INTELLIGENCE RELATED TO FATHER'S OCCUPATION 366
Choosing Employees by Mental and Physical Tests 388
CLASSIFICATION OF MEN, PRINCIPLES UNDERLYING THE 50
CLERKS (OFFICE), STANDARDIZED TEST FOR 248
CLINICAL PSYCHOLOGY, SYMPOSIUM ON 81
COACHPROOF TEST 172
Commercial Tests and How to Use Them 194, 387
COMPARATIVE VARIABILITY OF BOYS AND GIRLS 151
COMPARISON OF: COLORED AND WHITE CHILDREN, 277; CHILDREN IN GOOD AND POOR FARMING DISTRICTS, 283; TWO INTELLIGENCE SCALES, 310
CONFERENCE ON AMERICANIZATION IN INDUSTRIES 269
Creative Impulse in Industry 98
CROSS-OUT TESTS 138
CUED COMBINATION METHOD 183

Deficiency and Delinquency, an Interpretation of Mental Testing 386
DELINQUENTS AND ILLITERATES, INFORMATION TEST FOR, 374; AND DEPENDENTS, MENTAL RATING OF JUVENILE 291

EDUCATIONAL DIFFERENCES AMONG TRADESMEN 33
EFFICIENCY OF A GROUP SCALE OF INTELLIGENCE, 68; FOR SCHOLASTIC PROGNOSIS 381
Eugenics, Applied 96

FILING TEST 178
FLYING, TESTS OF APTITUDE FOR 103

GEOGRAPHICAL INFORMATION IN STUDENTS 186
GROUP EXAMINATION OF INTELLIGENCE INDEPENDENT OF LANGUAGE, 13 and 199; INTELLIGENCE SCALE FOR PRIMARY GRADES, 215; SCALE FOR SCHOLASTIC PROGNOSIS, 381; SCALE OF INTELLIGENCE, ITS PRACTICAL EFFICIENCY 68

INDIVIDUAL RECORDS COMPARED WITH STANDARD FILE 191
INDUSTRIES, CONFERENCE ON AMERICANIZATION IN 269
Industry, Creative Impulse in 98
INFORMATION TEST FOR DELINQUENTS AND ILLITERATES 374
INTELLIGENCE, EFFICIENCY OF A GROUP SCALE OF, 68; GROUP EXAMINATION OF, 13; OF SCHOOL CHILDREN RELATED TO FATHER'S OCCUPATION 366

Intelligence Scale, Absolute, 118; for Primary Grades, 215; Two Scales Compared 310
Intelligence Test, A Non-Language Group 199
Intellectual Development Despite Enforced Seclusion 167

Juvenile Dependents and Delinquents, Rating of 291

Kingdom of the Child 100

Learning Curve in Typewriting 252

Manual for Otis' Group Intelligence Scale 194
Mental Age of Adults 317

Nerve Control and How to Gain It 287
Notes on Bureau of Educational Research, Univ. of Ill., 196; Bureau of Salesmanship, Carnegie Inst. Tech., 197, 290 and 394; Children's Service Bureau, Youngstown, O., 289; Institute for Public Service, N. Y., 290; Iowa Child Welfare Research Station, 394; Mass. Society for Mental Hygiene, 396; National Research Council, 393; Society for American Fellowships in French Universities, 395; Training Course at the Athenaeum Press 396

Omnibus Test for Reasoning 189

Principles Underlying the Classification of Men 50
Psychology and Pedagogy of Anger 287

Rating of Juvenile Dependents and Delinquents 291
Religion and Culture, A Critical Survey 389

Scales of Intelligence, Comparison between Two, 310; Group Intelligence for Primary Grades 315
Standardized Test for Office Clerks 248
Symposium on Clinical Psychology 81

Telegraphers, Mental Tests for 110
Test, A Non-Language Group Intelligence, 199; for Mental Accuracy, 179; for Office Clerks, Standardized 248
Tests for Prospective Telegraphers, 110; of Aptitude for Flying 103
Trade Ability, Analysis of, 352; Test, A Written 358
Tradesmen, Educational Differences among 33
Typewriting, Learning Curve in 252

Variability of Boys and Girls 151

Winston Simplified Dictionary 194
Written Trade Test 358

INDEX OF AUTHORS

(The names of authors of original contributions are printed in SMALL CAPITALS.)

ARTHUR, GRACE 118

BILLS, M. A. 388
BINGHAM, W. V. 1
BLANCHARD, PHYLLIS 100, 287
Bruce, H. Addington 287

CARNEY, CHESTER S. 269
CHAPMAN, J. CROSBY 252, 358
Cody, Sherwin 194, 387

DOLL, EDGAR A. 317

FIELD, I. A. 287
FOSTER, JOSEPHINE CURTIS 167
FRASIER, GEORGE W. 151

GEISSLER, L. R. 98
GESELL, ARNOLD 81
GODDARD, HENRY H. 84
GOULD, MIRIAM C. 96

HAINES, THOMAS H. 291
Hall, G. Stanley 100
Heniger, Alice Minnie Herts 100
HENMON, V. A. C. 103

Johnson, Roswell H. 96

KARLSON, KARL J. 389
KELLEY, C. M. 172
KELLEY, TRUMAN L. 50
Kemble, William Fretz 383
KENAGY, H. C. 387
KITSON, HARRY DEXTER 310

Lewis, William D. 194
LOWELL, FRANCES 215

Marot, Helen 98
Miner, James Burt 386

Otis, Arthur S. 194

PARTLOW, W. D. 291
PAYNTER, RICHARD H., JR. 329
PINTNER, RUDOLF 33, 199
Popenoe, Paul 96
PRESSEY, L. W. 68, 138
PRESSEY, S. L., 68, 138, 277, 283, 366, 374, 381

RALSTON, RUTH 366
Richardson, Roy Franklin 287
ROBINSON, EDWARD S. 352

Schleiter, Frederick 389
SHIVELY, I. M. 374
Singer, Edgar A. 194
SPROWLS, J. W. 386
SWIFT, WALTER B. 156

TETER, G. F. 277
THOMAS, J. B. 283
THORNDIKE, EDWARD L. 13
THURSTONE, L. L. 110, 248
TOOPS, HERBERT A. 33, 358

WALLIN, J. E. WALLACE 87
WELLS, F. L. 172
WOODROW, HERBERT 118